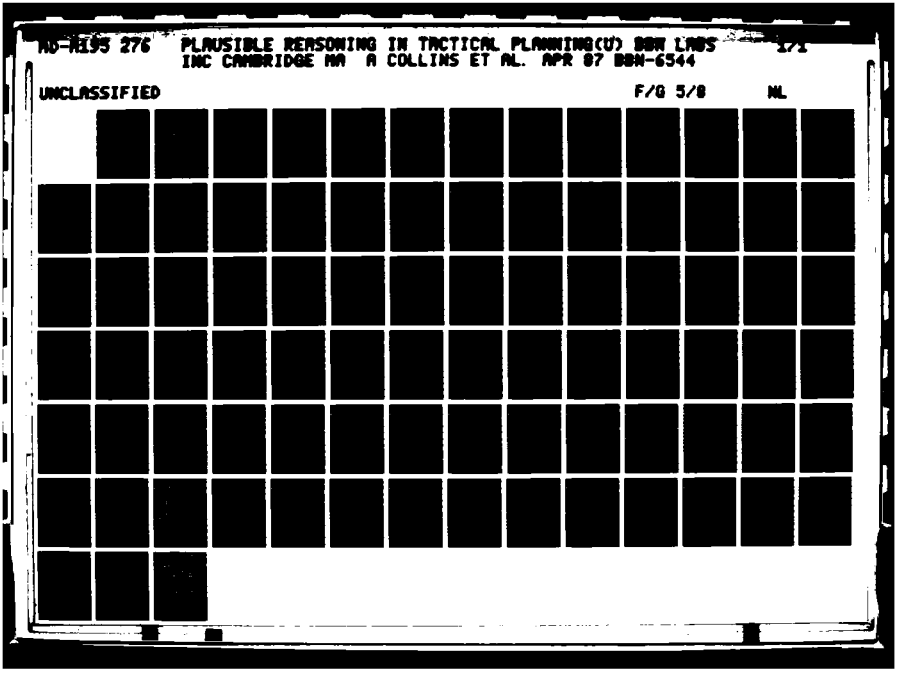
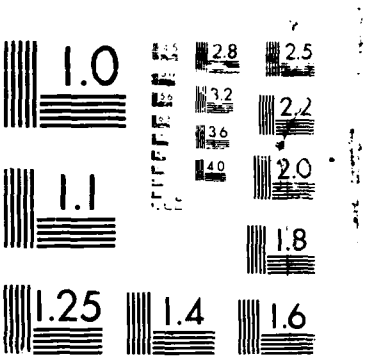

AD-A195 276 PLAUSIBLE REASONING IN TACTICAL PLANNING(U) BBN LABS INC CAMBRIDGE MA A COLLINS ET AL. APR 87 BBN-6544 1/1
UNCLASSIFIED F/G 5/8 NL



BBN Laboratories Incorporated

A Subsidiary of Bolt Beranek and Newman Inc.

Report No. 6544

Plausible Reasoning in Tactical Planning

Allan Collins, Mark Burstein, and Ryszard Michalski

May 1987

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 6544	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Plausible Reasoning in Tactical Planning		5. TYPE OF REPORT & PERIOD COVERED Annual Interim Report Oct. 1985 - Sept. 1986 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Allan Collins, Mark Burstein, and Ryszard Michalski		8. CONTRACT OR GRANT NUMBER(s) MDA903-85-C-0411
9. PERFORMING ORGANIZATION NAME AND ADDRESS BBN Laboratories Incorporated 10 Moulton Street Cambridge, MA 02238		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE April 1987 13. NUMBER OF PAGES 85
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Contracting officer's representative was Judith Orasanu.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

reasoning
artificial intelligence
similarity
analogy

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report includes two papers prepared during the year on a contract. The first paper details a formal theory of human plausible reasoning, and the second paper is an analysis of the literature on similarity and analogy.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Plausible Reasoning in Tactical Planning

Executive Summary

During the first year of the contract our work centered in four areas:

1. We developed a theory of human plausible reasoning which is detailed in a paper entitled "the Logic of Plausible Reasoning: A Core Theory" by Collins and Michalski which has been accepted by Cognitive Science, and is included in this report. The theory was based on detailed analyses of human reasoning about geography occurring in natural contexts such as tutoring and answering questions. The core theory constructed characterizes people's plausible deductions, inductions, and analogies in terms of Michalski's variable valued logic calculus.

2. We analyzed transcripts from videotapes of military planners considering where to deploy their troops in Europe against a particular massing of Red troops. We subsequently interviewed one of the military planners, Col. Fred Kulik, at length about how he thinks about different kinds of terrain and obstacles in assessing where an enemy might move their troops. We analyzed these protocols in terms of the theory of plausible reasoning detailed in the paper described above.

3. We built a preliminary computer model embodying the theory of human plausible reasoning outlined in the paper by Collins & Michalski. Our goal is to use the simulation as a means of testing and referring the theory. This requires developing appropriate memory organization and search techniques to support this style of inference, finding ways to estimate similarity in specific contests, and investigating ways of combining sometimes contradictory conclusions reached when inferences of different types are used to answer questions.

4. We wrote a paper analyzing the psychological literature on similarity and analogy entitled "A Framework for a Theory of Mapping" by Collins and Burstein. The paper analyzes the different kind of entities that are related by analogy and similarity mappings, the different contexts and tasks that give rise to mappings, and the set of issues and different resolutions to these issues that have been proposed in the literature. This paper is also included as part of this annual report.

THE LOGIC OF PLAUSIBLE REASONING: A CORE THEORY

Allan Collins
Bolt Beranek and Newman Inc.
Cambridge, MA

Ryszard Michalski
Computer Science Department
University of Illinois
Urbana, Illinois

TABLE OF CONTENTS

1. BACKGROUND FOR THE THEORY

The goal of our research on plausible reasoning is to develop a formal system based on Michalski's variable-valued logic calculus (1980, 1983) that characterizes different patterns of plausible inference humans use in reasoning about the world (Polya, 1958; Collins, 1978a). Our work attempts to formalize the plausible inferences that frequently occur in people's responses to questions for which they do not have ready answers (Carbonell & Collins, 1973; Collins, 1978a,b; Collins, Warnock, Aiello, & Miller, 1975). In this sense it is a major departure from formal logic, which represents normative theories of reasoning. Being descriptively based, it includes a variety of inference patterns that do not occur in formal logic-based theories. The central goals of the theory are to discover recurring general patterns of plausible inferences and to determine the parameters affecting the certainty of these inferences.

In order to analyze human plausible reasoning, Collins (1978b) collected a large number of people's answers to everyday questions, some from teaching dialogues and some from asking difficult questions to four subjects. These answers have the following characteristics:

1. There are usually several different inference patterns used to answer any question.

2. The same inference patterns recur in many different answers.

3. People weigh different evidence that bears on their conclusion.

4. People are more or less certain about their conclusion depending on the certainty of their information (either from some outside source or from memory), the certainty of the inference patterns and associated parameters used, and on whether different patterns lead to the same or opposite conclusions.

The analysis of the answers attempts to account for the reasoning and the conclusions drawn in terms of a taxonomy of plausible inference patterns. As will be evident, this is an inferential analysis. To use Chomsky's (1965) felicitous terms, we are trying to construct a deep structure theory from the surface structure traces of the reasoning process.

We will illustrate some of the characteristics of people's answers, as well as some of the inference patterns formulated in the theory with several transcripts. The first transcript comes from a teaching dialogue on South American geography (Carbonell & Collins, 1973) (T stands for teacher and S for student):

T. There is some jungle in here (points to Venezuela) but this breaks into a savanna around the Orinoco (points to the Llanos in Venezuela and Colombia).

S. Oh right, is that where they grow the coffee up there?

T. I don't think that the savanna is used for growing coffee. The trouble is the savanna has a rainy season and you can't count on rain in general. But I don't know. This area around Sao Paulo (in Brazil) is coffee region, and it is sort of getting into the savanna region there.

In the protocol the teacher went through the following reasoning. Initially, the teacher made a hedged "no" response to the question for two reasons. First, the teacher knew that coffee growing depends on a number of factors (e.g., rainfall, temperature, soil, and terrain), and that savannas do not have the correct value for growing coffee on at least one of those factors (i.e., reliable rainfall). In the theory this is an instance of the inference pattern called a *derivation from a mutual implication*. Second, the teacher did not know that the Llanos was used for growing coffee, which he implicitly took as evidence against its being a coffee region. The inference takes the form "I would know the Llanos produces coffee if it did, and I don't know it, so probably it does not." This is called a *lack-of-knowledge inference* (Collins et al., 1975; Gentner & Collins, 1982). This inference pattern is based on knowledge about one's own knowledge and hence is a meta-knowledge inference.

Then the teacher backed off his initial negative response, because he found positive evidence. In particular, he thought the Brazilian savanna might overlap the coffee growing region in Brazil around Sao Paulo, and therefore might produce coffee. If the Brazilian savanna produces coffee, then by functional analogy (called a

similarity transform in our theory) the Llanos might. Hence, the teacher ended up saying "I don't know," even though his original conclusion was correct.

The teacher's answer exhibits a number of the important aspects of human plausible reasoning. In general, a number of inference patterns are used together to derive an answer. Some of these are inference chains where the premise of one inference draws on the conclusion of another inference. In other cases the inference patterns are triggered by independent sources of evidence. When there are different sources of evidence, the subject weighs them together to determine a conclusion and the strength of belief in it.

It is also apparent in this protocol how different pieces of information are found over time. What appears to happen is that the subject launches a search for relevant information (Quillian, 1968; Collins & Loftus, 1975). As relevant pieces of information are found (or are found to be missing), they trigger particular inferences. Which inference pattern is applied is determined by the relation between the information found and the question asked. For the question about growing coffee in the Llanos, if the respondent knew that savannas are in general good for growing coffee, that would trigger a deductive inference. If the respondent knew of a similar savanna somewhere that produced coffee, that would trigger an analogical inference. The search for information is such that the most accessible information is found first, as by a marker passing or spreading activation algorithm (Charniak, 1982; Quillian, 1968).

In the protocol, the more accessible information about the unreliable rainfall in savannas was found before the less accessible information about the coffee growing region in Brazil and its relation to the Brazilian savanna. The order of finding information reflects its decreasing accessibility as activation spreads through a semantic network (Quillian, 1968). Relevant information is found by autonomous search processes, and the particular information found determines what inferences are triggered.

The next protocol illustrates a plausible deduction, called a *specialization transform* in the theory (Q stands for questioner and R for respondent):

Q. Is Uruguay in the Andes Mountains?

R. I get mixed up on a lot of South American countries (pause). I'm not even sure. I forget where Uruguay is in South America. It's a good guess to say that it's in the Andes Mountains because a lot of the countries are.

The respondent knew that the Andes are in most South American countries (7 out of 9 of the Spanish speaking countries). Since Uruguay is a fairly typical South American country, he guesses that the Andes may be there too. He is wrong, but the conclusion was quite plausible. This example illustrates a *specialization transform* and two of the certainty parameters associated with it : *frequency* (he knows the Andes are in most countries), and *typicality* (Uruguay is a typical South American country).

The third protocol illustrates another kind of plausible deduction, called a *derivation from mutual implication* in the theory:

Q. Do you think they might grow rice in Florida?

R. Yeah, I guess they could, if there were an adequate fresh water supply. Certainly a nice, big, warm, flat area.

The respondent knew that whether a place can grow rice depends on a number of factors. He also knew that Florida had the correct values on at least two of these factors (warm temperatures and flat terrain). He therefore inferred that Florida could grow rice if it had the correct value on the other factor he thought of (i.e., adequate fresh water). He may or may not have been aware that rice growing also depends on fertile soil, but he did not mention it here. Florida in fact does not produce rice in any substantial amount, probably because the soil is not adequate. This protocol shows how people make plausible inferences based on their approximate knowledge about what depends on what, and how the certainty of such inferences is a function of the degree of dependency between the variable in question (rice) and the known variables (i.e. terrain, climate, water).

The fourth protocol from a teaching dialogue illustrates a functional analogy, called the *similarity transform* in the theory:

S. Is the Chaco the cattle country? I know the cattle country is down there (referring to Argentina).

T. I think it's more sheep country. It's like western Texas, so in some sense I guess it's cattle country. The cattle were originally in the Pampas, but not so much anymore.

As in the first protocol, the respondent is making a number of plausible inferences in answering this question, some of which lead to different conclusions. First, he thinks that the Chaco is used for sheep raising, but there is some uncertainty about the information retrieved, which leads to a hedged response. This supports an implicit *lack-of-knowledge inference* (a meta-knowledge inference), that takes the form "I don't know that it's cattle country, and I would know if it were (e.g., I know about sheep), so it probably is not cattle country." But then the teacher noted a similarity between the Chaco and western Texas, presumably in terms of the functional determinants of cattle raising (e.g., climate, vegetation, terrain). This led him to a very hedged affirmative response, based on a *similarity transform*. Finally the teacher alluded to the fact that the Pampas is the place in Argentina known for cattle, and the place the student most likely was thinking of. This argues against the Chaco having cattle based on another meta-knowledge inference, a *confusability inference* (Collins, 1978b): "The Chaco is confusable with the Pampas and the Pampas has cattle, so the fact that there are cattle in Argentina cannot be taken as evidence for cattle in the Chaco." In answering this question, then, two patterns of plausible inference led to a negative conclusion and one to a positive conclusion.

The fifth protocol illustrates both a *similarity* and a *dissimilarity transform*, and more importantly, the distinction between inferences based on overall similarity and those based on similarity with respect to the functional determinants of the property in question.

Q. Can a goose quack?

R. No, a goose – well, its like a duck, but its not a duck.

It can honk, but to say it can quack. No, I think its vocal cords are built differently. They have a beak and everything, but no, it can't quack.

The *similarity transform* shows up in the phrases, "it's like a duck" and "They have a beak and everything" as well as the initial uncertainty about the negative conclusion. It takes the form, "A duck quacks and goose is like a duck with respect to most features, so maybe a goose quacks". The certainty of the inference depends on the degree of similarity between ducks and geese.

But then two lines of negative inference led the respondent to a negative conclusion. First there is a lack-of-knowledge inference implicit in the statement "It can honk, but to say it can quack." She knew about geese honking but not about their quacking. Therefore, she thought she would know about geese quacking, if in fact they did quack.

The second line of negative inference (apparently found after she started answering) is the dissimilarity inference evident when she says, "I think its vocal cords are built differently". The dissimilarity inference takes the form "Ducks quack, geese are dissimilar to ducks with respect to vocal cords, and vocal cords determine the sound an animal makes, so probably geese do not quack". This inference was enough to lead her to a strong "no". Of course she knew nothing about the vocal cords of ducks and geese, because they don't have any. She was probably thinking of the difference in the length of their necks. Our own hypothesis is that longer necks resonate at lower frequencies and hence honking can be thought of as deep quacking.

These five examples illustrate a number of aspects of human plausible reasoning as it occurs in common discourse. They show how people bring different pieces of knowledge to bear on a question and how these pieces sometimes lead to the same conclusion and sometimes to different conclusions. Often knowledge is found after the respondent has started answering, so that the certainty of the answer seems to change in midstream. The examples also show how people's approximate functional knowledge of what depends on what often comes to play in different inferences such as deductions and analogies. Therefore these dependencies are a central part of the

core theory we have developed. We will return to these examples to illustrate how the formal rules we have developed can be used to characterize different plausible inferences seen in these examples.

In our development of the theory to date we have not tried to characterize all the different types of plausible inferences that occur in the protocols. In particular we have not formalized the spatial and meta-knowledge inferences shown above. This project presents a core system centered around the plausible deductions, analogies, and inductions, seen most frequently in the protocols. In future work we plan to extend this core system to encompass the other patterns of inference, such as spatial and meta-knowledge inferences (Collins, 1978 a,b).

2. ASSUMPTIONS UNDERLYING THE THEORY

The theory assumes that a large part of human knowledge is represented in structures, we call *dynamic hierarchies*, that are interconnected by *traces*. Each hierarchy represents knowledge about a class of concepts arranged in a tree structure according to some viewpoint. Traces represent paths linking nodes in different hierarchies that record beliefs about the world. These beliefs can be recorded by our senses or derived by inference. The theory presented here shows that certain types of plausible inferences can be viewed simply as *perturbations* of traces in the knowledge structures.

The hierarchies are *dynamic* in that they are always being updated, modified or expanded. In the *core* theory described here we distinguish between two basic kinds of hierarchies, *type*- and *part*-hierarchies (Collins and Quillian, 1972). A *type*-hierarchy (also called an *abstraction* or *is-a* hierarchy) is organized by the *type* relation holding between connected nodes, or more precisely, between concepts represented by the nodes. A *part*-hierarchy is organized by the *part-of* relation holding between connected nodes. Any given node may be a member of more than one hierarchy. Each such hierarchy characterizes the node from a different viewpoint.

Nodes of a hierarchy may represent classes (e.g., flowers), individuals (e.g., a specific flower) or manifestations of individuals (e.g., a specific flower at a given moment). For the purpose of the theory, manifestations are treated just like individuals or classes.

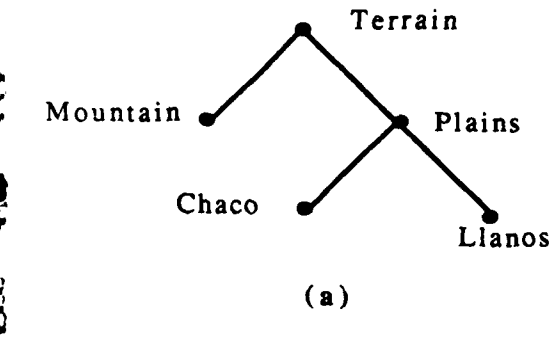
Figure 1 shows examples of type- and part-hierarchies. In the first four examples (1a,b,c,d), the Llanos is viewed from four different perspectives. These perspectives are organizing principles of the hierarchies (Bobrow and Winograd, 1977). The type-hierarchy in figure 1a is organized according to the type of terrain. The type of terrain can be mountainous, plateau, hilly, or plain, etc. The Llanos is characterized as a type of plain, like the Chaco. The type-hierarchy in figure 1b is organized according to the geographical land type. It characterizes the Llanos as a type of savanna, which is one of the major land types that geographers divide the world into, including rain forests, deserts, steppes, Mediterranean climates, mid-latitude forests, etc. The part-hierarchy in figure 1c is organized according to

regions in South America: the Andes, Amazon Jungle, Llanos, Guiana Highlands, and their subregions in different countries. The part-hierarchy in Figure 1d represents South America broken down into countries and the subregions within each.

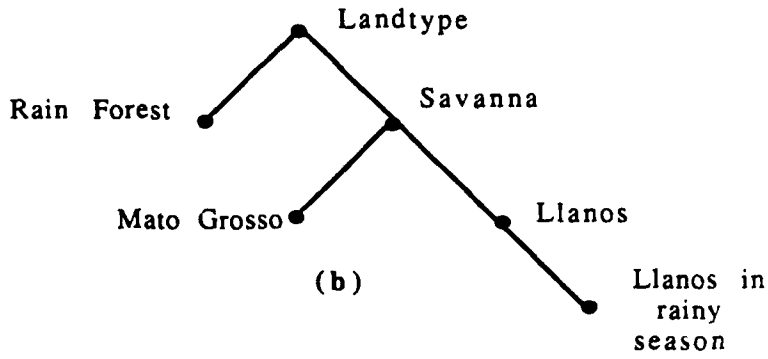
Insert Figure 1 here

The other three examples in Figure 1 are designed to illustrate how different descriptors also are represented in hierarchies. Among colors there are green and red. Among reds there are scarlet and burgundy, and among scarlets there are bright scarlet and perhaps dull scarlet, etc. Color is a one-place descriptor applying to objects, but feeling emotion is a two place descriptor where X (a person) feels the emotion toward Y (any concept). In the emotion hierarchy there are many types of emotions, among them love and hate, and there are different kinds of love, such as romance, affection, motherly love, etc. In the weight hierarchy there are different kinds of weight, such as human weight which in turn might be divided into birth weight and adult weight. For birth weight one might think of 1 lb. as a minimum, 15 lbs as a maximum, and 7 lbs as the norm. For the purposes of the theory these can be thought of as different values of birth weight, just as red and green are different values of color. These examples are not meant to show how people represent such concepts, but to give an idea as to how the hierarchies can represent different kinds of information.

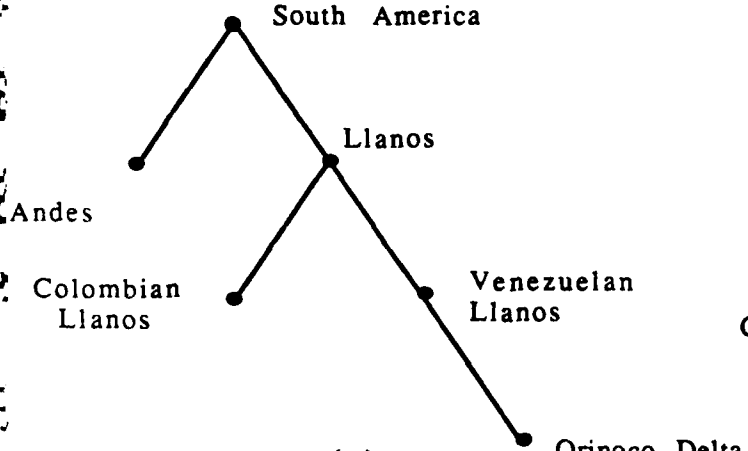
As mentioned above, traces represent recordings of information within the hierarchies. They are paths connecting the nodes of two or more hierarchies that represent *beliefs* about the world. Figure 2 shows examples of traces representing the beliefs that there are daffodils and roses in England, and that John's eyes are blue. The traces can have annotations describing their origin, their frequency of use, the certainty of belief in their correctness, and other information. The links denoting the type and part relation in generalization hierarchies can also be viewed as traces, but for the purpose of theory we will distinguish them from other traces. The knowledge organization described above includes various elements of semantic network structure (Carbonell & Collins, 1973; Collins & Quillian, 1972; Quillian, 1968) and frame structure (Bobrow & Winograd, 1977; Minsky, 1975; Schank & Abelson, 1977; Winograd, 1975).



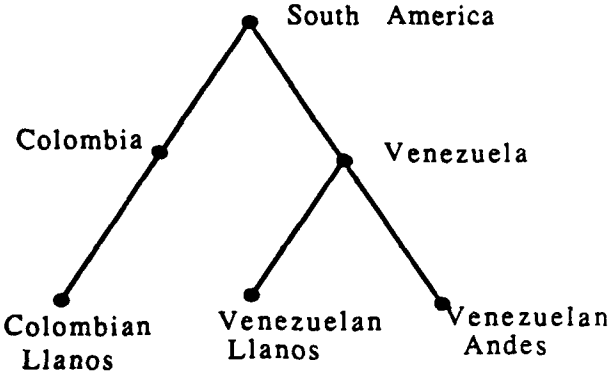
(a)



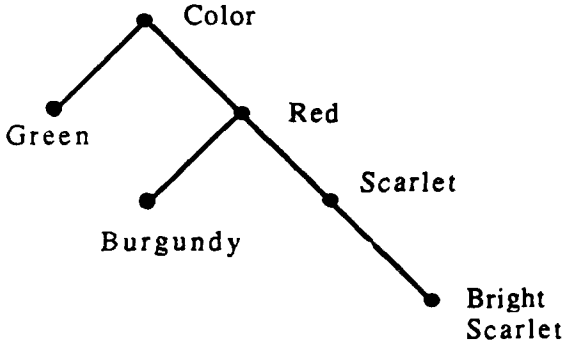
(b)



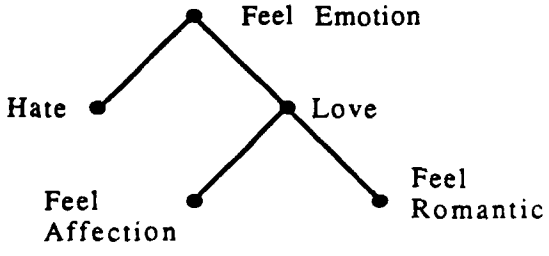
(c)



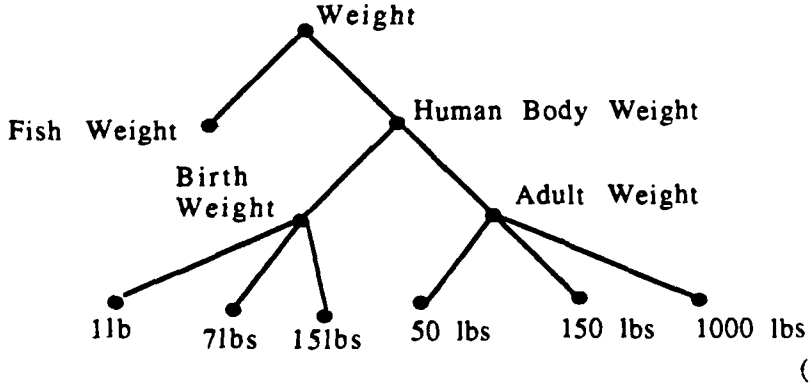
(d)



(e)



(f)



(g)

Figure 1. Examples of hierarchies.

Insert Figure 2 here

Let us explain some of the elements of annotations of a trace. By the *origin* of a trace we mean the information specifying whether the trace is a recording of a sense observation, an assertion obtained from a source of information (e.g., another person), or a statement derived through inference. Frequency of use or importance (Carborall & Collins, 1973, Collins & Quillian 1972, Collins & Loftus 1975) represents the ease of traversing a particular link, or the accessibility of one concept from another. Certainty of belief is discussed in detail in the next section.

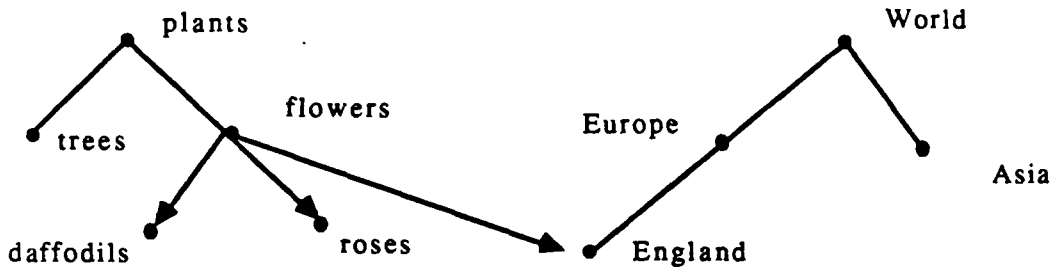
A trace may be a recording of information about one's beliefs, or denote the *applicability* relation between the nodes of different hierarchies. The applicability relation between a node A and a node B states that node A can be used as a *descriptor* of node B., i.e., that A can be used to characterize node B. We write such a relation as a term

A(B)

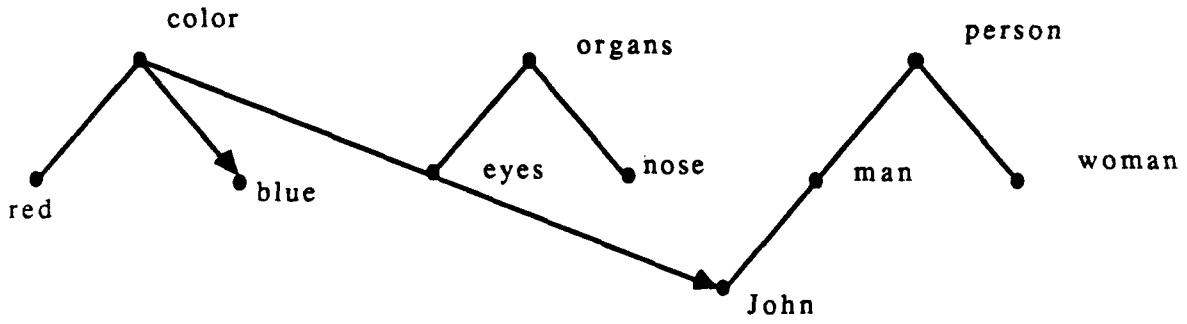
For example, the node "color" in hierarchy 1e applies as a descriptor to node "eyes" of hierarchy 1h. This is denoted as "color(eyes)." The node "eyes" can in turn be applied as descriptor to the node, say, John, in some hierarchy describing people. To express both relations we would write:

color(eyes(John))

A term A(B) can *take a value* only from the set of subnodes of A, i.e., the descendants of the node A in the hierarchy. The set of subnodes which can actually be a value of term A(B) is called the *domain* of term A(B). Applying a descriptor to an *argument* (node or a sequence of nodes) A produces a specific value characterizing the argument. This implies that only non-terminal nodes of a hierarchy can be descriptors. For example, to state that the color of the eyes of John is blue, a trace would be created that links John, color and blue as shown in Figure 2. To express this formally, we write:



flowers (England) = {daffodils, roses. . .}



color (eyes(John)) = blue

Figure 2. Examples of two traces on statements

color(eyes(John))=blue

In the theory such an expression is called a *statement*.

The applicability relation observes an important property. If it has been observed that A in a type-hierarchy is *applicable* to B in a type-hierarchy, then we can infer that A is applicable to any subnode of B, and that any supernode of A is applicable to B. For example, assume that the node "eyes" applies to "person". One can infer that also "organ" applies to "person" and that "eyes" applies to "woman." Part-hierarchies, for the most part, follow the same rules as type-hierarchies with some restrictions, such as the fact that a descriptor applicable to one node may not always apply to a subnode (e.g. capital applies to states but not to cities).

It is important to mention at this point that the applicability relation is learned like any other relation. This relation does not act as a "selection restriction" assumed by some linguists. Its violation is not considered to be a semantic anomaly, but rather as a new information to be made consistent with the existing knowledge structures. For example, when one hears that "an idea is green," then usually one tries to make sense of it rather than reject it as an anomalous expression.

Figure 3 illustrates the fact that the hierarchies are partial orderings, and can be differentiated or collapsed as appropriate for the purpose of drawing plausible inferences. At a fairly early age children think of animals as coming in different types: dogs, cats, fish, birds, etc. They don't differentiate them much more than that. When they get to school they may learn there are different basic types of animals, such as fish, birds, reptiles, mammals, and amphibians, and that dogs are cats are types of mammals. Still later in biology this hierarchy might be differentiated much more finely as in Figure 3c. But the early links are never lost; they are in fact used all the time in reasoning about the world. For the purpose of the theory, therefore, any hierarchy can be collapsed or differentiated as long as the partial orderings in the hierarchy are maintained.

Insert Figure 3 here

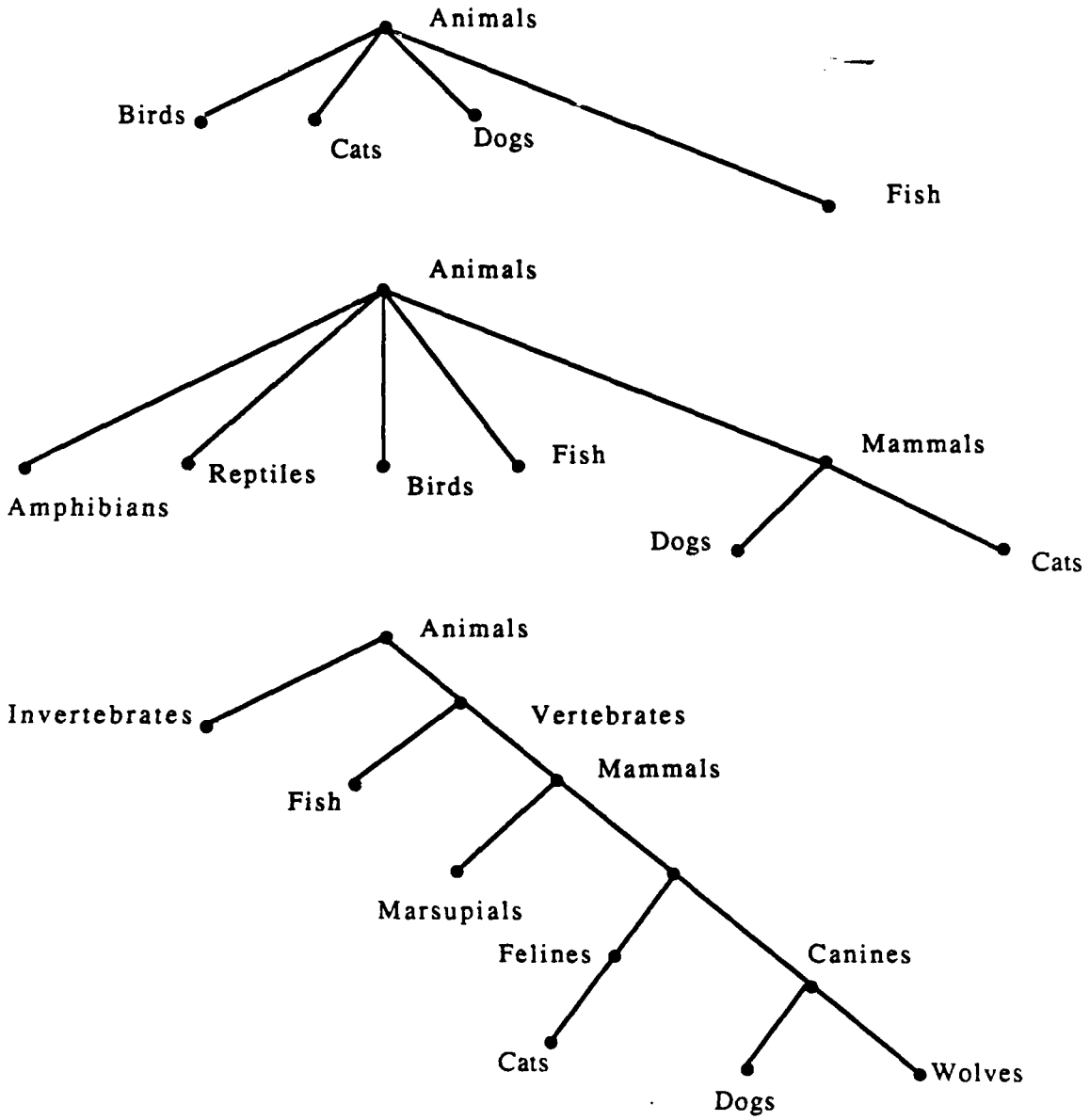


Figure 3: Differentiation of Hierarchies

Table 1 shows hypothetical frame structures for a few concepts in someone's memory (Collins & Quillian, 1972; Collins et al, 1975). These examples are not meant to provide a detailed analysis of how concepts are represented, but rather to illustrate how the statements shown in later examples can be constructed from a memory structure. In the example, type and part relations form the basis for hierarchical structures such as those shown in Figures 1 and 3. Flowers are represented as a type of plant coming in at least four varieties (i.e. roses, etc.), having various parts, various colors, and growing in all countries. Each descriptor (i.e. type/of, types, parts, color, countries) might be further specified as to how it relates to the concept flower (e.g., type/of is a biological class, colors are surface features of the petals, countries are places where flowers are grown, etc). Daffodils, which are a particular type of flower, provide further specification for each of the variables in the concept of flowers. That is, they have petals and a stem, they come in yellow and perhaps other colors, and they are grown in at least England and the United States. The frame for red is shown to illustrate how a color concept points back to various objects which it describes. Finally let us stress that we have not concerned ourselves with exactly how concepts are represented, but rather we have assumed they are represented in a structure similar to these examples.

Insert Table 1 here

Any node in a hierarchy can potentially be a *descriptor* for a node in another hierarchy. For example, if flower is in a hierarchy of things and England is in a hierarchy of places, flower-type might be a descriptor for England. This produces a *statement* of the form:

(1) flower-type (England)={daffodils, roses,....}

In (1) flower-type is a descriptor, England is an argument, flower-type (England) is a term, and daffodils and roses are references for the term. The brackets and dots indicate that daffodils and roses are not assumed to be a complete set, although the person may not know other flowers of England. Any descriptor, as a node in a hierarchy, can be further differentiated. For example, flowers can be differentiated between naturally-growing flowers vs. flowers grown in greenhouses, or between flowers sold vs. flowers grown, etc. People make finer or less fine discriminations

Table 1

Hypothetical Frames in a Person's Memory

flower

type/of =(plant)

types ={rose, daffodil, peony, bougainvillea ...}

parts ={petals, stem ...}

colors ={pink, yellow, white, red ...}

countries ={all countries}

daffodil

type/of =(flower)

parts ={petals, stem ...}

colors ={yellow ...}

countries ={England, United States ...}

red

type/of =(color)

types ={scarlet, burgundy ...}

flowers ={roses, tulips ...}

vehicles ={fire engines, London buses ...}

depending on their knowledge and purposes, and a theory of plausible reasoning must accommodate these different degrees of discrimination.

Whether a particular descriptor applies to any argument depends on what knowledge the person has. For example, it is not clear what red-type (England) might mean because one probably doesn't have knowledge in one's data base about the color of England (though one might interpret the term as the color of any part of England, such as the Union Jack and London buses).

Examples (2) to (8) below illustrate how different descriptors apply to different concepts:

(2) England-part (daffodil)={Southern England...}

(3) daffodil-part (England)={petals, stem...}

(4) country-type (daffodils)={temperate countries...}

(5) daffodil-type (England)={yellow daffodils...}

(6) England-type (daffodils)={England in the spring}

(7) love-type (John, Mary)={affection...}

(8) give-type (John, Mary, scarf)={gift-giving...}

Examples (2) and (3) illustrate statements based on part hierarchies. In (2) the descriptor selects the part of England where daffodils occur. In (3) the descriptor selects the parts of daffodils that occur in England; presumably daffodil parts in England are the same as daffodil parts anywhere in the world (though perhaps Martian daffodils are quite different). In (4) country-type applied to daffodils selects the types of countries that have daffodils (i.e., temperate countries). Statement (4) could have specified the particular countries (e.g. England, France) that have daffodils, since hierarchies can be collapsed as long as a partial order is maintained. In (5) daffodil-type applied to England selects the different daffodil types found in England, of which only one type is stored (i.e., yellow daffodils), though there may be others. In (6) we show that when you take an instance like England and look at its subtypes you get a manifestation, in this case the manifestation(s) that have daffodils. Finally, (7) and (8)

illustrate multiple place predicates describing John's love of Mary, and John's giving a scarf to Mary as a gift rather than loaning it or giving it away to get rid of it. These examples show how different terms are evaluated within the theory.

These examples illustrate the most important assumptions we are making about how human memory is organized and accessed for the purposes of making plausible inferences. Further descriptions of our underlying assumptions about human memory are given in earlier papers (Carbonell & Collins, 1973; Collins & Loftus, 1975; Collins & Quillian, 1972; Collins, Warnock, Aiello & Miller, 1975).

3. PRIMITIVES IN THE CORE SYSTEM,

In the core system we have developed there is a set of primitives and a set of basic inference rules. In this section we describe the primitives in the system, consisting of basic expressions, operators, and certainty parameters.

Table 2 shows the basic elements in the core system. *Arguments* can be any node in a hierarchy, or a function of one or more nodes such as Fido's master or the flag of England. *Descriptors* apply to arguments, and together they form a *term*, such as breed (Fido). The *reference* for a term can be either a definite set of values such as collie, or brown and white, or an indefinite set of values such as brown plus other colors (or possibly no other colors).

Insert Table 2 here

Statements consist of a term on the left of an equals sign and a reference on the right, together with a set of certainty parameters. Expressions (1) through (8) above were all statements, without the certainty parameters specified. The operator statements shown below in Table 3 are a special class of statements. The certainty parameters can be thought of as approximate numbers ranging between 0 and 1, but we have represented them as verbal descriptions. In the example shown, γ refers to how certain one is the statement is true, and ϕ to the frequency that if something is a bird it can fly. These certainty parameters are all listed in Table 4, to be discussed later.

The last two types of expressions represent functional dependencies between different variables. *Dependencies between terms* represent the functional relationship between two terms, such as between the average temperature of a place and the latitude of the place. The dependency can be annotated to different degrees: it can be unmarked meaning there exists some functional relation the two, it can be marked with + or − indicating a monotonic increasing or decreasing relation, or it can be further specified to any degree (e.g., a V-shaped function with 3 values specified). For example, if one thinks that average temperature of a place in January varies between about 85° at the equator and −30° at the North Pole and + 30° at the South

Table 2

Elements of Expressions

arguments a_1, a_2, $f(a_1)$
e.g., Fido, collie, fido's master

descriptors d_1, d_2
e.g., breed, color

terms $d_1(a_1)$, $d_2(a_2)$, $d_2(d_1(a_1))$
e.g., breed (Fido), color (collie), color (breed (Fido))

references r_1, (r_2,r_3), $\{r_2 \ldots\}$
e.g., collie, brown and white, brown plus other colors

statements $d_1(a_1)=r_1$: γ, ϕ
e.g., means of locomotion (bird)={flying...}: certain, high frequency

dependencies between terms $d_1(a_1)<--->d_2(f(a_1))$: α, β, γ
e.g., latitude (place) <----> average temperature (place): moderate, moderate, certain

implications between statements $d_1(a_1)=r_1<===>d_2(f(a_1))=r_2$: α, β, γ
e.g., grain (place)={rice...} <===> rainfall (place)=heavy: high, low, certain

Pole, this relation can be represented as a V-shaped function with values (-90°, 30°), (0°, 85°) and (90°, -30°), where the first coordinate is latitude and the second temperature. The α and β parameters specify the degree of constraint in the dependency from latitude to temperature and from temperature to latitude, respectively. In the latitude-temperature example the degree of constraint is moderate in both directions, as is discussed later.

Implications between statements relate particular values of functions such as the latitude-temperature function above (e.g., latitude (place) = equator <=> average temperature (place) = hot). The example shown in the table relates the grain of a place being rice to the rainfall of the place being heavy (e.g., >40 in/year). Knowing a place produces rice predicts that it will have heavy rainfall quite strongly, so that α is high (though there are exceptions like Egypt where rice is grown by irrigation). However the fact that the rainfall of a place is heavy (e.g., Oregon) only weakly predicts that rice is grown, so β is low. In general mutual implications between statements will be asymmetric in this way.

Table 3 illustrates the four operators in the core system and the kinds of statements they occur in. The generalization and specialization operators go up and down in a hierarchy, while the similarity and dissimilarity operators go between nodes at the same level in a hierarchy. Associated with the GEN and SPEC operators there is a typicality parameter τ (Rosch, 1975; Smith & Medin, 1982), and with the SIM and DIS operators there is a similarity parameter σ. There is also a dominance parameter δ associated with GEN and SPEC statements that specifies what proportion of the superset, the subset actually comprises. Finally all the statements involving operators have a certainty parameter γ associated with them.

Insert Table 3 here

Typicality and similarity are always computed in some context which is denoted by the CX variable. The first variable in the CX denotes a node in the argument hierarchy specifying the range of arguments over which typicality or similarity are computed. For GEN and SPEC this is always the superset specified in the statement (e.g., for chicken=SPEC (barnyard fowl), barnyard fowl is the superset over which

Table 3

Operators

Generalization a'=GEN(a) in CX(a,D): γ, τ, δ

e.g., bird=GEN(chicken) in CX (birds, physical features): certain, atypical, low dominance

Specialization a'=SPEC(a) in CX(a',D): γ, τ, δ

e.g., chicken=SPEC(barnyard fowl) in CX (barnyard fowl, food cost): certain, typical, moderate dominance

Similarity a'=SIM(a) in CX(A,D): γ, σ

e.g., ducks=SIM(geese) in CX(birds, all features): certain, highly similar

Dissimilarity a'=DIS(a) in CX(A,D): γ, σ

e.g., ducks=DIS(geese) in CX(birds, neck length): certain, fairly dissimilar

typicality is computed, but for SIM and DIS it is the basic level category (Rosch 1975; Smith & Medin, 1982) to which the two arguments belong that is the basis for computing similarity. Hence the similarity of ducks and geese would normally be computed in the context of birds, which is their basic level category.

The second variable in the CX specifies the set of descriptors to be used in comparing the two nodes with respect to typicality or similarity. For example, one can evaluate how typical chickens are as birds with respect to their physical features, with respect to all their features, or with respect to some particular feature such as the cost of feeding them. Similarity and dissimilarity can also be computed with respect to different features. As we discussed with respect to the fifth protocol shown earlier, ducks and geese are quite similar when compared on all their features, but they are dissimilar in neck length (which is relevant to determining the sound they make). The procedure for computing typicality and similarity is described below.

Table 4 lists the certainty parameters we have identified so far that affect the certainty of different plausible inferences. We will describe each of these parameters in terms of the examples given above. The description is meant to specify our best hypothesis about how people might compute these parameters.

Insert Table 4 here

The α and β parameters can best be introduced in terms of the example: grain(place)={rice...}<===>rainfall(place)=heavy. As we said, α would be high in such case if a person thinks that most places that grow rice have heavy rainfall (say > 40 inches per year), whereas β would be low if he or she thinks there are many places with heavy rainfall, that don't produce rice. We can construct a hypothetical table that represents this view in terms of a small sample of places and the frequencies with which they have heavy rainfall and produce rice:

	Rice	No Rice	Total
Heavy Rainfall	8	8	16
No Heavy Rainfall	2	20	22
Total	10	28	38

Table 4

Certainty Parameters

α Likelihood that the right-hand side of a dependency or implication is in a particular range given that the left-hand side is in a particular range.

β Likelihood that the left-hand side of a dependency or implication is in a particular range given that the right-hand side is in a particular range.

γ Degree of certainty that a statement is true (i.e., degree of belief).

τ Degree of typicality of a subset within a set (e.g., robin is a typical bird and ostrich is an an atypical bird).

σ Degree of similarity of one set to another set.

ϕ Frequency of the reference in the domain of the descriptor (e.g., above 90% of birds fly).

δ Dominance of a subset in a set (e.g., chickens are not dominant among birds, but are dominant among barnyard fowl).

Given this table α is simply the conditional probability that a rice-producing place has heavy rainfall, in this case 8 of 10 or .8 and β is the conditional probability that a place with heavy rainfall produces rice, in this case 8 of 16 or .5. We don't think that people actually construct such tables though they may consider a small number of cases in computing rough estimates of α and β, as they do in using the availability heuristic (Tversky & Kahneman, 1973).

The α and β parameters for mutual dependencies can be constructed by an extension of the procedure for mutual implications. Suppose one considers the relationship of rainfall and grain growing as before, but instead as a mutual dependency (i.e., grain (place) <--> rainfall (place). For simplicity we can present the same hypothetical table in revised form:

	Rice	Wheat	Corn	Total
Heavy Rainfall	8	6	2	16
Light Rainfall	2	14	6	22
Total	10	20	8	38

Then α reflects the degree to which you can predict whether a place has heavy or light rainfall, given the predominant grain grown in the place, which is quite high (i.e., the prediction is correct in 28 or 38 cases or .7 assuming an optimal guessing strategy). Similarly, β reflects the degree to which you can predict whether they grow rice, wheat, or corn, given the amount of rainfall (i.e., the prediction is correct in 22 of 38 cases or .6, assuming an optimal strategy of guessing wheat for light rainfall and rice for heavy rainfall). This example makes evident the fact that the α and β parameters reflect the way the dependency partitions the known cases in the world.

The γ parameter in Table 3 reflects the certainty or subjective likelihood with which a person believes any expression is true. γ can reflect different possible sources of uncertainty. One source arises when people retrieve a fact from memory and are uncertain they may be making a memory confusion. Another basis for uncertainty arises when they doubt the source from which they got the information. Finally, if a piece of information derives from a plausible inference, there will be uncertainty as to whether the conclusion is correct, and this uncertainty will propagate to inferences dependent on it. All these sources of uncertainty are represented by the γ parameter.

Typicality (τ) and similarity (σ) can be thought of as the same parameter: in the case of typicality it is computed between a subset and its superset, and in the case of similarity it is computed between two subsets. We assume that any set (or concept) is represented as a bundle of features (Collins & Quillian, 1972), and the τ and σ parameters are computed by comparing the two concepts with respect to those features specified by the descriptor variable in the context CX. For example, "chicken" might be compared to "bird" with respect to size or with respect to all its physical features to determine its typicality. For a continuous variable like size, typicality or similarity is determined by computing how close (normalized between 0 and 1) the two values are in the distribution of sizes for the class specified by the context CX (e.g. birds). For discrete variables like "ability to fly", the two concepts either match or not (assigned either 1 or 0). Typicality or similarity are based on the average score for all the features compared, weighted for their criteriality or importance (Carbonell & Collins, 1973; Collins & Quillian, 1972).

Frequency (ϕ) and dominance (δ) reflect different ratios that affect the certainty of plausible inferences in systematic ways. Frequency reflects the proportion of members of the argument set that can be characterized by the reference specified. It reflects what "Some" or "All" reflect in logic (e.g., "Some men have arms"), but as a continuous variable between 0 and 1. For the statement "means-of-locomotion (birds)={flying...}," is the proportion of birds that fly to the total of all birds. The dominance of a subset within a set (δ) applies only to generalization and specialization statements. It reflects the proportion of members of the set that are members of the subset specified in the statement. For example, chickens constitute a high proportion of barnyard fowl, but not of birds in general.

This completes our summary of the primitives in the system. We will now describe the different plausible inference forms in the core system.

4. TRANSFORMS ON STATEMENTS

The simplest class of inferences in the core theory are called transforms on statements. If a person believes some statement, such as that the flowers growing in England include daffodils and roses [i.e., flower-type(England)={daffodils, roses...}], there are eight transforms of the statement that allow plausible conclusions to be drawn. These eight transforms can be thought of as perturbations of the statement either with respect to the argument hierarchy (starting from England) or the reference hierarchy (starting from daffodils and roses). The argument-based transforms move up (using GEN), down (using SPEC), or sideways (using SIM or DIS) in the argument hierarchy. Similarly the reference-based transforms move up, down, or sideways in the reference hierarchy. Thus each of these transforms is a perturbation in one of the two hierarchies.

Let us illustrate the eight transforms on statements in terms of hierarchies for England and roses. Figure 4 shows a part hierarchy for England and a type hierarchy for roses and daffodils that someone might have. If the person believes that, "flower-type(England)={daffodils, roses...}," then Table 5 shows eight conclusions that the person might plausibly draw.

Insert Figure 4 and Table 5 here

The first GEN inference is that Europe as a whole grows daffodils and roses. This may not be true: Daffodils and roses may be a peculiarity of England, but it is at least plausible that daffodils and roses are widespread throughout Europe. Similarly, for the SPEC relation it is a plausible inference that the county of Surrey in southern England grows roses and daffodils. There is an implicit context (CX) in GEN and SPEC transforms, that will be discussed later.

The SIM and DIS inferences are also made in some context. In the case of the argument-based transforms the context might be "countries of the world with respect to the variable climate." Holland is quite similar to England with respect to climate, while Brazil is quite dissimilar. The variables over which the comparison is made may be few or many, but people will make the comparison with respect to those variables

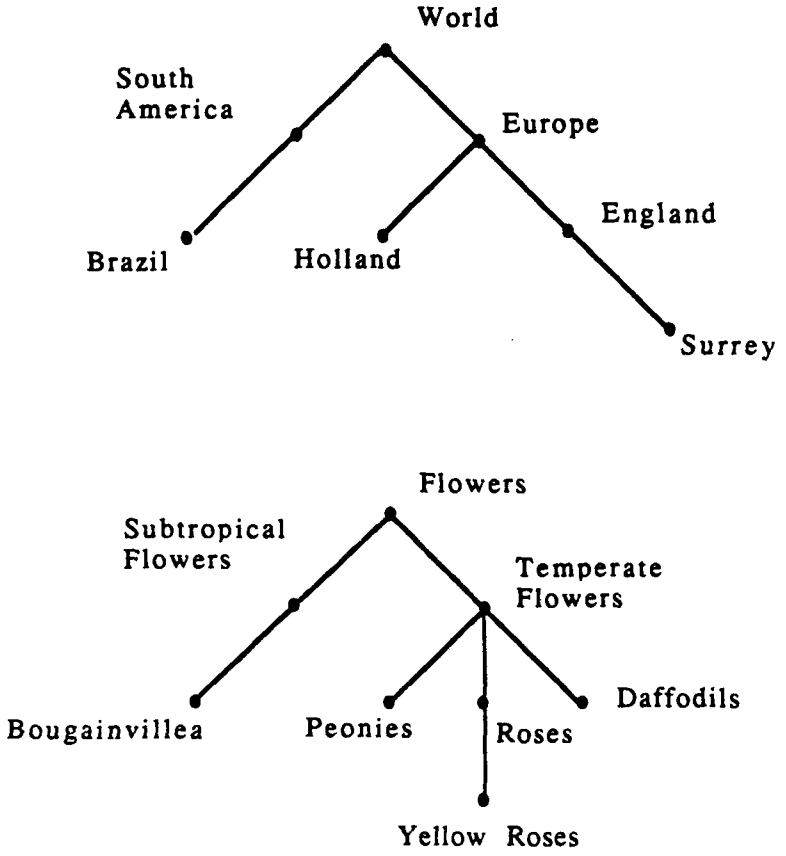


Figure 4. Part Hierarchy for England and Type Hierarchy for Roses.

Table 5

Eight Transforms on the Statement
"flower-type(England)={daffodils, roses...}"

Argument-based Transforms

(1)	GEN	flower-type(Europe)={daffodils, roses...}
(2)	SPEC	flower-type(Surrey)={daffodils, roses...}
(3)	SIM	flower-type(Holland)={daffodils, roses...}
(4)	DIS	flower-type(Brazil)≠{daffodils, roses...}

Reference-based Transforms

(5)	GEN	flower-type(England)={temperate flowers...}
(6)	SPEC	flower-type(England)={yellow roses...}
(7)	SIM	flower-type(England)={peonies...}
(8)	DIS	flower-type(England)≠{bougainvillea...}

that they think are most relevant to the question (e.g., whether they grow daffodils in Holland). That is, they base their inference on whatever mutual dependency most constrains the descriptor in question. In this case the flowers grown in a place depend highly on the climate of the place, but hardly at all on the longitude of the place. Therefore climate is a reasonable variable on which to make the comparison. We will refer to this issue later when we talk about how different parameters affect the certainty of any statement transform.

The reference transforms are perhaps easiest to understand if you substitute a fictional place like Ruritania for England, because other inferences are not invoked so easily. If one believes they grow daffodils and roses in Ruritania, then one might infer they grow temperate flowers in general there, and yellow roses in particular. It is also reasonable that they grow peonies there, since they are similar to roses and daffodils as to the climates they grow in. But bougainvillea grows in more tropical climates, so it is unlikely to grow in Ruritania (Ruritania is, after all, a small little kingdom and unlikely to encompass different climates--this is a supporting inference). These examples should give a feel for how the transforms on statements are made.

4.1 Certainty Parameters Affecting Transforms on Statements

In this section we will discuss how different certainty parameters affect the various transforms shown in Table 5.

Typicality. Typicality (τ) affects the certainty of any GEN or SPEC transform as shown in Table 6. In argument-based transforms the more typical the subset is of the set in the argument hierarchy, the more certain the inference. For example, in Table 5 inference (1) is more certain the more typical England is as part of Europe.

Insert Table 6 here

In making plausible inferences people compute typicality with respect to those variables, such as climate, that they think flower growing depends on. Thus, if Surrey is thought to have a typical climate for England, and climate is thought to predict the types of flowers grown in a place, then the inference is more certain.

Table 6

Effects of Different Parameters on Statement Transforms

Transforms in Table 5		Parameters					Target Node
		τ	σ	α	ϕ	δ	
Argument-Based	1 GEN	+	0	+	+	+	Europe
	2 SPEC	+	0	+	+	+	Surrey
	3 SIM	0	+	+	+	0	Holland
	4 DIS	0	−	+	−	0	Brazil
Reference-Based	5 GEN	+	0	+	+	+	Tropical Plants
	6 SPEC	+	0	+	+	+	Yellow Roses
	7 SIM	0	+	+	+	0	Peonies
	8 DIS	0	−	+	−	0	Bougainvillea

Note: As the value of the parameter increases, a + means it has a positive effect on the certainty of the inference and a − means it has a negative effect on the certainty of the inference.

This example reveals the mutual dependency implicit in any statement transform. The mutual dependency relates the set of variables on which the typicality or similarity judgment is made (e.g., climate or all variables) to the descriptor in question (e.g., flower-type). If the typicality judgment is made considering all variables (as when we said Surrey is a typical English county), the transform will be inherently less certain because of the weak dependency between most variables and any descriptor such as flower-type. Therefore, if you know that Surrey is typical of England in general, it leads to a less certain inference than if you know Surrey is typical of England with respect to climate.

In a reference-based transform typicality works the same way, except that it is computed with respect to the subset and its superset in the reference hierarchy. In inference (5) in Table 5, the greater the typicality of daffodils and roses as temperate plants, the more certain the inference. Similarly in the inference (6), the greater the typicality of yellow roses as roses, the more certain the inference. Pink roses are more typical than yellow roses, and so they are even more likely to be found in England (or Ruritania for that matter). Again the inference is more certain if typicality is measured with respect to the climate in which the flowers are grown.

Similarity. Degree of similarity (σ) affects the certainty of any SIM or DIS inference as shown in Table 6. Like typicality, similarity can be computed over all variables or over a subset of variables (e.g., climate) that are particularly relevant.

Degree of similarity increases the certainty of SIM inferences and decreases the certainty of DIS inferences, as would be expected. In Table 5, therefore the inference (3) that Holland has daffodils and roses is more certain the more similar Holland is to England with respect to climate or whatever variables one thinks flowers are related to. The inference (4) that Brazil does not have roses and daffodils is more certain the less similar Brazil is to England. The inference (7) that England has peonies is more certain, the greater the similarity of peonies to both daffodils and roses. The inference (8) that England does not have bougainvillea is more certain, the less similar bougainvillea is to daffodils and roses. More particularly bougainvillea is dissimilar in that it tends to grow in warmer climates than daffodils and roses.

Mutual Dependency. Every transform on a statement involves an implicit mutual

dependency. The inference is always more certain the greater the dependency ($\propto$) between the variables on which typicality or similarity are measured and the variable in question as shown in Table 6. If climate were the variable used for measuring typicality and similarity, the argument-based transforms would be more certain the more the climate of a place constrains the flowers grown in the place. The mutual dependency is slightly different for reference-based transforms. They would be more certain, the more the climate where flowers grow constrains the places where flowers grow.

Frequency. The frequency (ϕ) of the reference set within the domain of the argument affects the certainty of all eight inferences, as shown in Table 6. For an instance, e.g. England, frequency with respect to the argument set only makes sense if you think of England as a set of small parts (say 10 miles square) and count the frequency of parts that have daffodils and roses vs. those that do not. The more frequent daffodils and roses are in the parts of England, then all but the DIS inferences are more certain. For example, roses and daffodils are more likely to occur in Holland or Surrey if they are very frequent in England. The two DIS inferences go in the opposite direction. For example, the less frequent are daffodils and roses in England, the more likely bougainvillea will be found there (though this is a very weak inference).

Dominance. Dominance (δ) affects GEN and SPEC inferences as is shown in Table 6. In all cases, the greater the dominance of the subset, the more certain the inference. For example, for (2) if Surrey comprised most of England it would be a more certain inference that it has daffodils and roses, than if it is a very small area in England. Similarly for (6) if yellow roses were the most dominant kind of roses, they would be more likely found in England than if they are a rare type of rose.

4.2 Formal Representation of Transforms on Statements

Table 7 shows the formal representations we have developed for each of the eight transforms on statements in terms of the variable-valued notation of Michalski (1983). Most of the examples shown are from protocols we have collected (Collins, 1978b), some of which appear in the first section of this paper. We will briefly describe each of the examples.

Insert Table 7 here

We can illustrate an argument-based transform or GEN with the inference that if chickens have gizzards, then birds in general may have gizzards. The first premise, represents the belief that chickens have gizzards: presumably almost all chickens have gizzards so the frequency (ϕ) and the certainty (σ) arer high. The second premise represents the belief that chickens are birds, and that they are typical with respect to their biological characteristics. As we pointed out earlier, the subset dominance (δ) of chickens among birds is low. The third premise states that the internal organs of a bird depend highly on the biological characteristics of the bird. The conclusion that birds have gizzards is fairly certain given the high values of the critical variables.

The argument-based transform on SPEC is illustrated by an example from the beginning of the paper where the respondent inferred that the Andes might be in Uruguay. The respondent believed that the Andes are in most South American countries, so frequency (ϕ) was moderately high. With respect to the second premise, Uruguay is a typical South American country, which increases the likelihood that the Andes would be found there. But its low subset dominance (δ) in terms of the proportion of South America that Uruguay comprises makes the inference less likely. With respect to the third premise, the fact that Uruguay is typical of South American countries in general only weakly predicts that it will include the Andes mountains. Altogether, the inference is fairly uncertain given the moderate frequency and the low subset dominance of Uruguay.

We can illustrate the argument-based transform on SIM with the Chaco protocol from the beginning of the paper, where the respondent inferred that the Chaco might produce cattle given that west Texas did. In the first premise, frequency (ϕ), which reflects the degree to which different parts of west Texas have cattle, is high, which makes the inference more likely. The second premise asserts that the Chaco is a least moderately similar to west Texas in vegetation (or whatever variables the respondent had in mind). The third premise relates vegetation of a region to its livestock, which is a strong relation, given that cattle will usually be raised where the vegetation will support them. The fourth premise merely establishes the fact that west Texas and

Table 7

Formal Representations of Statement Transforms

(1) Argument-based transform on GEN

d(a)=r: γ_1 , ϕ
a'=GEN(a) in CX (a',D(a')): τ, γ_2, δ
D(a') <-----> d(a'): α , γ_3

---

d(a')=r: $\gamma = f(\gamma_1, \phi, \tau, \gamma_2, \delta, \alpha, \gamma_3)$

Internal organ (chicken) = {gizzard ...}: γ_1 =high, ϕ =high
Birds = GEN (chicken) in CX (bird, biological characteristics(birds)):
τ =high, γ_2=high, δ =low
Biological characteristics (birds) <-----> internal organs (birds):
α =high, γ_3=high

Internal organs (birds) = {gizzard ...}: γ=high

(2) Argument-based transform on SPEC

d(a)=r: γ_1, ϕ
a'=SPEC(a) in CX (a, D(a)): τ , γ_2, δ
D(a) <-----> d(a): α , γ_3

---

d(a')=r: $\gamma = f(\gamma_1, \phi, \tau, \gamma_2, \delta, \alpha, \gamma_3)$

Mountains(S.A. country) = {Andes ...}: γ_1=high, ϕ=high,
Uruguay=SPEC(S.A. country) in CX(S.A. country, characteristics(S.A. country)):
τ =high, γ_2 =high, δ =low
Characteristics (S.A. country) <-----> mountains (S.A. country):
α =moderate, γ_3=high

Mountains (Uruguay) = {Andes ...}: γ=moderate

(3) Argument-based transform on SIM

d(a) = r: γ_1, ϕ
a' = SIM (a) in CX (A, D(A)): σ, γ_2
D(A) <----> d(A): α, γ_3
a, a' = SPEC (A): γ_4, γ_5

---

d(a') = r: γ = f (γ_1, ϕ, σ, γ_2, α, γ_3, γ_4 γ_5)

Livestock (West Texas) = {cattle ...}: γ_1=high, ϕ =high
Chaco = SIM (West Texas) in CX (region, vegetation(region)):
σ =moderate, γ_2 =moderate
Vegetation (region) <----> livestock (region): α=high, γ_3=high
West Texas, Chaco = SPEC (region): γ_4=high, γ_5=high

Livestock (Chaco) = {cattle ...}: γ=moderate

(4) Argument-based transform on DIS

d(a) = r: γ_1, ϕ
a' = DIS(a) in CX(A, D(A)): σ, γ_2
D(A) <----> d(A): α, γ_3
a, a' = SPEC(A): γ_4, γ_5

---

d(a')$\neq$ r: γ= f(γ_1, ϕ, σ, γ_2, α, γ_3, γ_4, γ_5)

Sound (duck) = (quack): =high, =high
Goose = DIS (duck) in CX(bird, vocal cords (bird)):
=low, =moderate
Vocal cords (bird) <----> sound (bird): =high, =low
Duck, goose = SPEC (bird): =high, =high

Sound (goose)$\neq$quack: γ=low

(5) Reference-based transform on GEN

d(a) = {r ...}: γ_1, ϕ
r' = GEN (r) in CX(d, D(d)): τ, γ_2, δ
D(d) <-----> A(d): α , γ_3
a = SPEC (A): γ_4

---

d(a) = {r'...}: $\gamma = f(\gamma_1, \phi, \tau, \gamma_2, \delta, \alpha, \gamma_3, \gamma_4)$

Agricultural product (Honduras) = {bananas ...}:
γ_1=unknown, ϕ=high,
Tropical fruits = GEN (bananas) in CX(agricultural products,
climate(agricultural products)): τ=high, γ_2 =high, δ =low
Climate (agricultural products) <-----> Place (agricultural products):
α=high, γ_3 =high
Honduras = SPEC (place): γ_4 =high

Agricultural products (Honduras)={tropical fruits...}: γ=moderate

(6) Reference-based transform on SPEC

d(a) = {r ...}: γ_1, ϕ
r' = SPEC (r) in CX(d, D(d)): τ , γ_2, δ
D(d) <-----> A(d): α , γ_3
a = SPEC (A): γ_4

---

d(a) = {r'...}: γ = f (γ_1, ϕ, τ, γ_2, δ, α, γ_3, γ_4)

Minerals (South Africa) = {diamonds. γ_1 = high , ϕ = high
Industrial diamonds=SPEC(diamonds) in CX(minerals, characteristics(minerals)):
τ =high, γ_2 =high, δ =high
Characteristics(minerals) <-----> Place (minerals):
α =moderate, γ_3 =high
South Africa = SPEC (place): γ_4=high

Minerals (South Africa) = {industrial diamonds ...}: γ =high

(7) Reference-based transform on SIM

d(a) = {r...}: γ_1 , ϕ
r' = SIM (r) in CX(d, D(d)): σ , γ_2
D(d) <-----> A(d): α , γ_3
a = SPEC (A): γ_4

---

d(a) = {r'...}: $\gamma = f(\gamma_1, \phi, \sigma, \gamma_2, \alpha, \gamma_3, \gamma_4)$

Sound (wolf) = {howl...}: γ_1 =high, ϕ =high,
Bark = SIM (howl) in CX(sound, means of production(sound)):
σ =high, γ_2=high
Means of production (sound) <-----> animal (sound): α=high, γ_3 =high
Wolf = SPEC (animal): γ_4=high

Sound (wolf) = {bark...}: γ =moderate

(8) Reference-based transform on DIS

d(a) = {r...}: γ_1 , Φ
r' = DIS (r) in CX(d, D(d)): σ , γ_2
D(d) <-----> A(d): α , γ_3
a = SPEC(A): γ_4

---

d(a)$\neq$ {r'...}: $\gamma = f(\gamma_1, \Phi, \sigma, \gamma_2, \alpha, \gamma_3, \gamma_4)$

Color (Princess phones) = {white, pink, yellow...}: γ_1 =high, ϕ =high
Black = DIS (white & pink & yellow) in CX(color, hue(color)):
σ =low, γ_2 =high
Hue (color) <-----> object (color): α=low, γ_3 =high
Princess phone = SPEC (object): γ_4 =high

Color (Princess phones)$\neq${black...}: γ =moderate

Chaco are regions, in support of the second and third premises. The conclusion is only moderate in certainty, given our assumption of uncertainty about how similar the Chaco and west Texas are.

To illustrate the argument-based transform on DIS, we chose the example from the protocol shown earlier as to whether a goose quacks. The first premise reflects the respondent's belief that ducks quack, which was very certain. The second premise states the belief that ducks and geese are dissimilar in their vocal cords which the respondent must have been at least a bit uncertain about (hence the low certainty assigned to the statement). The third premise relates the sound a bird makes to its vocal cords, which also must have been an uncertain belief given that it is not true. The certainty of the conclusion that geese do not quack should have been fairly low (though other inferences led to the same conclusion in the actual protocol).

We have created an example to illustrate a reference-based transform on GEN, since there are none in the protocols. The first premise asserts that Honduras produces bananas among other things. Bananas are a fairly typical tropical fruit in terms of the climates where they are grown, as the second premise states. The third premise asserts that the climate appropriate for agricultural products constrains the places where they are grown fairly strongly. The conclusion follows with moderate certainty that Honduras produces tropical fruits in general, such as mangos and coconuts.

We also created the example of a referenced-based transform on SPEC. The first premise states that South Africa produces diamonds. Industrial diamonds are a kind of low quality diamond (used in drills) and they must be fairly dominant (δ) among diamonds given their low quality, though they are not particularly typical of what we think of as diamonds. Here is a case where high dominance compensates for low typicality. The third premise is somewhat irrelevant since the typicality is low. But the inference is quite certain given the high dominance of industrial diamonds among diamonds.

The example of a reference-based transform on SIM is drawn from a protocol where the respondent, when asked whether wolves could bark, inferred they probably could (Collins, 1978b). One of his inferences derived from the fact that he knew

wolves could howl, with both high frequency and certainty. He also thought that barking was similar to howling in terms of the way the sound is produced (a howl, as it were, is a sustained bark). Further the animals that make a particular sound depend on the means of production of the sound, as the third premise states. It follows then with at least moderate certainty that a wolf can bark.

The example of a reference-based transform in DIS is from a protocol where the respondent was asked if there are black princess telephones (Collins, 1978b). The respondent could remember seeing white, pink and yellow princess phones, as the first premise states. Here the frequency (ϕ) of these colors among those she had seen seemed quite high, which counts against the possibility of black princess phones. The second premise reflects the fact that black is quite dissimilar to those colors in terms of hue. The third premise states that the object associated with a particular color depends weakly ($\propto$ is low) on the hue of that color (i.e., knowing the hue only somewhat constrains the object). The conclusion that princess phones are not black is uncertain given the low $\propto$ in the third premise.

5. OTHER INFERENCES IN THE CORE THEORY

There are a number of other inference patterns in the core theory we have developed. In this section we will give the formal representation for each of the other inference patterns together with an example of each.

Table 8 shows that two types of *derivation from mutual implication* that occurred in the protocols shown at the beginning of the paper. The positive derivation illustrates how multiple conditions were ANDed together (i.e., a warm climate, heavy rainfall, and flat terrain) as predictors of rice growing. The belief that Florida has all three leads to a prediction that rice will be grown there. In the actual protocol the respondent was unsure about rainfall in Florida, and so concluded that rice would be grown if there was enough rain (i.e., Rainfall(Florida) = heavy <===> Product(Florida) = {rice...}). This is a slight variation on the positive derivation that can be represented as follows:

$$d_1(a) = r_1 \wedge d_2(a) = r_2 \Longleftrightarrow d_3(a) = r_3 : \alpha, \gamma_1$$
$$d_1(a') = r_1 : \phi, \gamma_2$$
$$\underline{a' = SPEC(a) : \gamma_3}$$
$$d_2(a') = r_2 \Longleftrightarrow d_3(a') = r_3 : = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$$

Insert Table 8 here

The negative derivation illustrates the fact that if any of the variables on one side of a mutual implication that are ANDed together do not have the appropriate values, then you can conclude that the variable on the other side does not have the value assumed in the mutual implication. In the example, because the Llanos did not have reliable rainfall, the respondent concluded that the Llanos probably did not produce coffee. If variables are ORed together (e.g., either heavy rainfall or irrigation are needed for growing rice) a different pattern holds: having one or the other predicts rice is grown and having neither predicts no rice is grown.

Table 9 shows the equivalent representations for derivations from mutual

Table 8

Formal Representations of Derivations from Mutual Implication

Positive Derivation

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2 : \alpha, \gamma_1$

$d_1(a') = r_1 : \phi, \gamma_2$

$\underline{a' = SPEC(a) : \gamma_3}$

$d_2(a') = r_2 : \gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$

Climate(place) = warm $\wedge$ Rainfall(place) = heavy $\wedge$ Terrain(place) = flat $\Longleftrightarrow$
Product(place) = {rice...} : α = high, γ_1 = certain
Climate(Florida) = warm : ϕ_1 = moderately high, γ_2 = certain
Rainfall(Florida) = heavy : ϕ_2 = moderate, γ_3 = uncertain
Terrain(Florida) = flat : ϕ_3 = high, γ_4 = certain
$\underline{\text{Florida} = SPEC(place) : \gamma_5 = \text{certain}}$
Product(Florida) = {rice...} : γ = uncertain

Negative Derivation

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2 : \alpha, \gamma_1$

$d_1(a') \neq r_1 : \phi, \gamma_2$

$\underline{a' = SPEC(a) : \gamma_3}$

$d_2(a') \neq r_2 : \gamma = f(\alpha, \gamma_1, \phi, \gamma_2, \gamma_3)$

Rainfall(place) = reliable $\wedge$ climate(place) = subtropical $\Longleftrightarrow$
Product(place) = {coffee...} : α = moderate, γ_1 = certain
Rainfall(Llanos) $\neq$ reliable : ϕ = high, γ_2 = fairly certain
$\underline{\text{Llanos} = SPEC(place) : \gamma_3 = \text{certain}}$
Product(Llanos) $\neq$ {coffee...} : γ = fairly certain

dependencies. It is impossible to draw a negative conclusion from a mutual dependency, since it denotes how a whole range of values on one variable relates to a range of values on another variable. But the inference patterns are different for positive and negative dependencies, so we have separated them in the table.

Insert Table 9 here

The positive dependency represents the case where as one variable increases, the other variable also increases. In the formal analysis we have denoted the entire range of both variables by three values: high, medium, and low. When a positive dependency holds, if the values of the first variable is high, medium, or low, the value of the second variable will also be high, medium, or low, respectively. This is the weakest kind of derivation possible from a mutual dependency: In the example, if a person knows that the temperature of air predicts the water holding capacity of air, and he knows that temperature of the air outside is high, then he can infer that the air outside could hold a lot of moisture. People make this kind of weak inference very frequently in reasoning about such variables (Collins & Gentner, in press; Stevens & Collins, 1980).

The pattern for the negative dependency is reversed: if the value of one variable is high, the other is low, and vice versa. We have illustrated the derivation from a negative dependency in terms of a more precise dependency between two variables. If a person believes that the latitude of a place varies negatively (and linearly) with the temperature of the place, and also that the average temperature is near 85 degrees at the equator and 0 degrees at the poles, then he might conclude that a place like Lima, Peru, that is about 10 degrees from the equator, has an average temperature of about 75 degrees. People have both more and less precise notions of how variables interact, and we have tried to preserve flexibility within our representation for handling these different degrees of precision.

Table 10 shows two forms of a transitive inference, one based on mutual implication and the other based on mutual dependency. The example for mutual implication states that if a person believes an average temperature of 85 degrees implies a place is equatorial, and that if a place is equatorial it will tend to have high humidity, then he can infer that if the average temperature of a place is 85 degrees it

Table 9

Formal Representations of Derivations from Mutual Dependencies

Derivation from Positive Dependency

$d_1(a) <--\overset{+}{}--> d_2(a) : \alpha, \gamma_1$
$d_1(a') =$ high, medium, low $: \Phi, \gamma_2$
$a' =$ SPEC(a) $: \gamma_3$
$d_2(a') =$ high, medium, low $: \gamma = f(\alpha, \gamma_1, \Phi, \gamma_2, \gamma_3)$

Temperature(air) $<--\overset{+}{}-->$ Water holding capacity(air) $: \alpha =$ high, $\gamma_1 =$ certain
Temperature(air outside) = high $: \Phi =$ high, $\gamma_2 =$ certain
Air outside = SPEC(air) $: \gamma_3 =$ certain
Water holding capacity(air outside) = high $: \gamma =$ certain

Derivation from Negative Dependency

$d_1(a) <--\overset{-}{}--> d_2(a) : \alpha, \gamma_1$
$d_1(a') =$ high, medium, low $: \Phi, \gamma_2$
$a' =$ SPEC(a) $: \gamma_3$
$d_2(a') =$ low, medium, high $: \gamma = f(\alpha, \gamma_1, \Phi, \gamma_2, \gamma_3)$

Abs. Val. Latitude(place) $<--\overset{-}{}-->$ Aver. Temperature(place): linear;
$0°, 85°; 90°, 0°; \alpha =$ moderate, $\gamma_1 =$ certain
Abs. Val. Latitude(Lima Peru) $= 10° : \Phi =$ high, $\gamma_2 =$ fairly certain
Lima Peru = SPEC(place) $: \gamma_3 =$ certain
Aver. Temperature(Lima Peru) $= 75°: \gamma =$ moderately certain

will tend to have high humidity, and vice versa. This example illustrates the way people confuse causality and diagnosticity in their understanding. If one were to write the causal links for this example, it would probably go from equatorial latitude to high temperature to high humidity. But people do not systematically make a distinction between causal and diagnostic links, nor do they store things in such a systematic order. For example, they may know that equatorial places, such as jungles, have high humidity and not link it explicitly to their high temperature. Thus, the inference in this example derives a more direct link (temperature <==> humidity) from a less direct link (latitude <==> humidity). It also should be noted that the diagnostic link in the first implication (temperature => latitude) may be more constraining than the causal link (latitude => temperature). That is, there are probably more equatorial places where the average temperature is not 85 degrees (e.g. Ecuador), than places where the temperature is 85 degrees but are not equatorial.

Insert Table 10 here

The example for a transitivity inference on mutual dependency illustrates how people reason about economics (Salter, 1983). Salter asked subjects questions, such as what is the effect of an increase in interest rates on the inflation rate of a country. People gave him chains of inferences like the one shown: if interest rates increase, then growth in the money supply will decrease, and that in turn will cause the inflation rate to decrease (the latter is a positive relation). So an increase in interest rates will lead to a decrease in the inflation rate. This kind of reasoning is a major way that people construct new mutual inplications and dependencies.

Tables 11 and 12 show a set of transforms on mutual implications that follow the same pattern as the transforms on statements in the previous section. Table 11 shows four reference transforms that parallel the last four statement transforms shown in Tables 5 and 7. (In fact there is a quite direct equivalence, because any statement can be transformed into a mutual implication in the following way: Flowers (England) = {daffodils...} goes into type(place) = England <==> flowers(place) = {daffodils...}, or more generally, d(a) =r goes into type(A)=a <==> d(A) = r.) We have represented the three positive transforms (i.e. generalization, specialization, and similarity) in the rule at the top, with the three alternatives shown (GEN, SPEC, and SIM) where they occur

Table 10

Formal Representations of Transitivity Transforms

On Mutual Implication

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2 : \alpha_1, \beta_1, \gamma_1$

$\underline{d_2(a) = r_2 \Longleftrightarrow d_3(a) = r_3 : \alpha_2, \beta_2, \gamma_2}$

$d_1(a) = r_1 \Longleftrightarrow d_3(a) = r_3 : \alpha = f(\alpha_1, \alpha_2),\ \beta = f(\beta_1, \beta_2)\ \gamma = f(\gamma_1, \gamma_2)$

Aver. Temperature(place) = 85° <==> Latitude(place) = equatorial :
α_1 = high, β_1 = fairly high, γ_1 = certain
Latitude(place) = equatorial <==> Abs. humidity(place) = high :
$\underline{\alpha_2 = \text{high},\ \beta_2 = \text{moderate},\ \gamma_2 = \text{certain}}$
Aver. Temperature(place) = 85° <==> Abs. Humidity(place) = high :
α = high, β = low, γ = certain

On Mutual Dependency

$d_1(a) \longleftrightarrow d_2(a) : \alpha_1, \beta_1, \gamma_1$

$\underline{d_2(a) \longleftrightarrow d_3(a) : \alpha_2, \beta_2, \gamma_2}$

$d_1(a) \longleftrightarrow d_3(a) : \alpha = f(\alpha_1, \alpha_2),\ \beta = f(\beta_1, \beta_2),\ \gamma = f(\gamma_1, \gamma_2)$

Interest rates(country) <-?-> Money supply growth(country):
α_1 = high, β_1 = moderate, γ_1 = certain
Money supply growth(country) <+?-> Inflation rate(country):
$\underline{\alpha_2 = \text{high},\ \beta_2 = \text{high},\ \gamma_2 = \text{certain}}$
Interest rates(country) <-?-> Inflation rate (country):
α = high, β = low, γ = certain

in the rule. The typicality parameter (τ) is associated with the GEN and SPEC transforms, and the similarity parameter (σ) with the SIM transform. The example omits the certainty parameters for simplicity. In English the example states the following: given the belief that if a place is subtropical, it is likely to produce oranges, this implies that if a place is subtropical, it is likely to produce citrus fruits (a generalization), or naval oranges (a specialization), or grapefruit (a similarity transform). The dissimilarity transform at the bottom follows the same pattern: if you think that subtropical places produce oranges, and apples are dissimilar to oranges with respect to their growing conditions, then probably subtropical places do not produce apples.

Insert Table 11 here

Table 12 shows the corresponding four types (i.e., GEN, SPEC, SIM, and DIS) of argument transforms. These correspond to the first four statement transforms shown in Tables 5 and 7. We illustrate the four with a demographic example: if one believes that men who live in the tropics have a short life expectancy and that farmers are typical *of men in terms of* their demographic characteristics, then one can plausibly infer that farmers have a short life expectancy if they live in the tropics. Similarly one can infer that people in general and women (because they are similar to men in their demographic characteristics) have short life expectancy in the tropics. Finally, one might conclude that birds do not have a short life expectancy in the tropics, if one thinks they are dissimilar to men in their demographic characteristics.

Insert Table 12 here

Table 13 shows the corresponding positive transforms for mutual dependencies. We have illustrated these with another example from economics: if one believes that the business tax rate in a state negatively impacts the amount of investment in the state, then one might generalize this relationship to any governmental unit, or particularize it to Illinois, or conclude that it would also apply to Canadian provinces. There is really no negative transform based on dissimilarity that corresponds to these three positive transforms. For example, if one believes that communist countries are quite dissimilar from states in their economics, the most one can conclude is that

Table 11

Formal Representations of Reference Transforms on Mutual Implications

Positive Transforms

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2 : \alpha_1, \gamma_1$

{GEN }

r'_2 = {SPEC} r_2 in CX(d_2, $D(d_2)$) : $\{\overset{\tau}{\sigma}\}, \gamma_2$

{SIM }

$\underline{D(d_2) \longleftrightarrow A(d_2) : \alpha_2, \gamma_3}$

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r'_2 : \gamma = f(\alpha_1, \gamma_1 \{\overset{\tau}{\sigma}\} \gamma_2, \alpha_2\ \gamma_3)$

Climate(place) = subtropical <==> Fruit(place) = {oranges...}
{Citrus fruits} {GEN }
{Naval oranges} = {SPEC} (oranges) in CX (fruit, growing conditions(fruit))
{Grapefruit} {SIM }
Growing conditions(fruit) <--> Place(fruit)

{Citrus fruit...}
Climate(place) = subtropical <==> Fruit(place) = {Naval oranges...}
{Grapefruit...}

Negative Transform

$d_1(a) = r_1 \Longleftrightarrow d_2(a) = r_2 : \alpha_1, \gamma_1$

r'_2 = DIS r_2 in CX (d_2, $D(d_2)$) : σ , γ_2

$\underline{D(d_2) \longleftrightarrow A(d_2) : \alpha_2, \gamma_3}$

$d_1(a) = r_1 \Longleftrightarrow d_2(a) \neq r'_2 \ : \gamma = f(\alpha_1, \gamma_1, \sigma, \gamma_2, \alpha_2, \gamma_3)$

Climate(place) = subtropical <==> Fruit(place) = {oranges...}
Apples = DIS(oranges) in CX (fruit, growing conditions (fruit))
Growing conditions(fruit) <--> Place (fruit)
Climate(place) = subtropical <==> Fruit(place) ≠ {apple...}

Table 12

Formal Representations of Argument Transforms on Mutual Implications

Positive Transforms

$d_1(a) = \quad <==> d_2(a) = r_2 : \alpha_1, \gamma_1$

$a' = \begin{matrix}\{GEN\}\\ \{SPEC\}\\ \{SIM\}\end{matrix}$ (a) in CX (A, $d_3(A)$) : $\{\begin{matrix}\tau\\ \sigma\end{matrix}\}, \gamma_2$

$\underline{d_3(A) <---> d_2(A) : \alpha_2, \gamma_3}$

$d_1(a) = r_1 <===> d_2(a) = r_2 : \gamma = f(\alpha_1, \gamma_1, \{\begin{matrix}\tau\\ \sigma\end{matrix}\}, \gamma_2, \alpha_2, \gamma_3)$

Habitat(man) = tropics <==> Life expectancy (man) = short

$\begin{matrix}\{GEN\}\\ \{SPEC\}\\ \{SIM\}\end{matrix}$ (farmer)

Man = {SPEC} (person) in CX(people, demographic characteristics(people))

{SIM } (woman)

<u>Demographic characteristics(people) <--> life expectancy(people)</u>

$\begin{matrix}(farmer)\\ (person)\\ (woman)\end{matrix}$ Habitat = tropics <==> life expectancy $\begin{matrix}(farmer)\\ (person)\\ (woman)\end{matrix}$ = low

Negative Transforms

$d_1(a) = r_1 <==> d_2(a) = r_2 : \alpha_1, \gamma_1$

$a' = DIS(a)$ in $CX(A, d_3(A)) : \sigma, \gamma_2$

$\underline{d_3(A) <---> d_2(A) : \alpha_2, \gamma_3}$

$d_1(a') = \quad <==> d_2(a') = \gamma : f(\alpha_1, \gamma_1, \sigma, \gamma_2, \alpha_2, \gamma_3)$

Habitat(man) = tropics <==> life expectancy(man) = short

Man = DIS(bird) in CX(animals, demographic characteristics (animals))

<u>Demographic characteristics(animals) <--> life expectancy (animals)</u>

Habitat(birds) = tropics<==> life expectancy(birds) = low

there is no negative relation between the business tax rate (if there were one) and the amount of investment; that is to say, no conclusion can be drawn. In such a case we just omit the form from the theory, because the theory does not specify conclusions that cannot be drawn. Similarly, there can be no reference transforms on mutual dependencies, because they do not involve a reference term.

Insert Table 13 here

Table 13

Formal Representations of Argument Transforms on Mutual Dependencies

Positive Transforms

$d_1(a) <\overset{+}{-}> d_2(a) : \alpha_1, \gamma_1$

$a' = \begin{Bmatrix} \text{GEN} \\ \text{SPEC} \\ \text{SIM} \end{Bmatrix} (a) \text{ in CX } (A, d_3(A)) : \left\{ \frac{\tau}{\sigma} \right\}, \gamma_2$

$\underline{d_3(A) <--> d_2(A) : \alpha_2, \gamma_3}$

$d_1(a') <\overset{+}{-}> d_2(a') : \quad = f(\alpha_1, \gamma_1, \frac{\tau}{\sigma}, \gamma_2, \alpha_2, \gamma_3)$

Business tax rate (state) <$\overset{-}{-}$> Amount of investment (state))

$\begin{Bmatrix} \text{Government unit} \\ \text{Illinois} \\ \text{Province} \end{Bmatrix} = \begin{Bmatrix} \text{GEN} \\ \text{SPEC} \\ \text{SIM} \end{Bmatrix}$ (state) in CX(place, economics (place))

<u>Economics(place) <--> Amount of investment(place)</u>

(government unit) (government unit)

Business tax rate (Illinois) <$\overset{-}{-}$> Amount of investment (Illinois)

(province) (province)

6. CONCLUSION

The difficulty in constructing a theory of plausible reasoning from analyzing actual cases of human reasoning is that the theory is likely to be underconstrained. That is to say, there may be many cases where people could employ a particular reasoning pattern, but do not because of other constraints on its invocation. As it stands now, the only constraints we place on the invocation of any inference pattern is that its premises be satisfied and that its certainty parameters not drive the conclusion below some threshold level of certainty. But there may well be other factors that constrain the invocation of any inference pattern.

In order to test out the core theory, we plan to build a computer system incorporating the reasoning patterns derived from our analysis. We will then be able to see what inferences the system draws given different knowledge bases. We plan to evaluate the theory in a series of experiments comparing the system's reasoning to that of expert human reasoners. To do this we will ask expert human reasoners, working from well-specified, small knowledge bases to draw plausible conclusions from each knowledge base and to estimate the certainty of each conclusion. These experts will be asked to put aside, as best they can, other knowledge they may have about the domain.

At the same time we will run the system on each small knowledge base to see what plausible conclusions the system draws, and with how much certainty. For each knowledge base, then we will have three different classes of inference: conclusions both computer and experts draw, conclusions the computer draws but experts do not, and conclusions experts draw that the computer does not draw. The two non-overlapping lists require different kinds of refinement to the theory. Where the computer draws a conclusion experts do not, we will go to the experts to see if the conclusion seems at all plausible to them. If not, then the set of inference rules must be modified to prevent such implausible conclusions from being drawn. Where experts draw a conclusion that the computer does not, we will first have to ascertain if they are drawing upon information the computer does not have. If not, then new inference rules must be added to the system to produce the conclusions that the human experts drew. The modifications to the theory will be implemented in a new version of the system, and the whole process will recycle until a stable state is reached, where the system and expert reasoners draw the same conclusions from new knowledge bases.

7. ACKNOWLEDGEMENTS

This research was supported by the Army Research Institute under Contract No. MDA 903-85-C-0411, by the National Institute of Education under Contract No. HEW-NIE-400-80-0031 by the National Science Foundation under Grant No. DCR-84-06801, and by the Office of Naval Research under Grant No. N00014-82-K-0186.

8. REFERENCES

Bobrow, D.G. & Winograd, T. (1977). An overview of KRL, a knowledge representation language. Cognitive Science, 1, 3–46.

Carbonell, J.R. & Collins, A., (1973). Natural semantics in artificial intelligence. Proceedings of Third International Joint Conference on Artificial Intelligence. Stanford CA: Stanford University, 344–351.

Charniak, E. (1983). Passing markers: A theory of contextual influence in language comprehension. Cognitive Science, 7, 171–190.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA. MIT Press.

Collins, A. (1978a). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.) Theoretical Issues in Natural Language Processing II. Urbana, IL: University of Illinois.

Collins, A. (1978b). Human plausible reasoning. Cambridge, MA: Bolt Beranek and Newman Inc., Report No. 3810.

Collins, A. & Gentner D. (in press). How people construct mental models. In N. Quinn and D. Holland (Eds.) Cultural models in language and thought. Cambridge University Press, Cambridge, UK.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407–428.

Collins, A.M. & Quillian, M.R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), Organization of Memory. New York: Academic Press.

Collins, A., Warnock, E.H., Aiello, N. & Miller, M.L. (1975). Reasoning from Incomplete Knowledge, in D. Bobrow & A. Collins (Eds.). Representation & understanding: Studies in cognitive science. New York: Academic Press.

Gentner, D.G. & Collins, A. (1981). Studies of inference from lack of knowledge. Memory & Cognition, 9, 434–443.

Michalski, R.S. (1980). Pattern recognition as rule-guided inductive inference. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-2, 349-361.

Michalski, R.S. (1983). A theory and methodology of inductive learning. Artificial Intelligence, 20, 111-161.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill.

Polya, G. (1968). Patterns of plausible inference. Princeton, NJ: Princeton University Press.

Quillian, M.R. (1968). Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, MA: MIT Press.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.

Salter, W. (1983). Tacit theories of economics. In Proceedings of the Fifth Annual Conference of the Cognitive Science Society. Rochester, NY: University of Rochester.

Schank, R. & Abelson, R. (1977). Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Smith, E.E. & Medin, D.L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Stevens, A. & Collins, A. (1980). Multiple conceptual models of a complex system. In R. Snow, P. Federico, & W. Montague (Eds.), Aptitude, learning and instruction: Cognitive process analysis. Hilldale, NJ: Erlbaum.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Winograd, T. (1975). Frame representations and the declarative-procedural controversy. In D.G. Bobrow & A. Collins (Eds.). Representation and understanding. New York: Academic Press.

A Framework for a Theory of Mapping

Allan Collins and Mark Burstein

BBN Laboratories Inc.
Cambridge, MA 02238

To appear in S. Vosniadou & A. Ortony (Eds.)
Similarity and Analogical Reasoning.

The writing of this paper was supported by the Army Research Institute under Contract No. MDA 903-85-C-0411.

1. INTRODUCTION

The literature on similarity, analogy, and metaphor ranges over many different kinds of mappings. Some of the disagreements arise because researchers are talking about different kinds of mappings or the different contexts in which mappings are made. Our goal is to clarify the issues being addressed and the critical distinctions that need to be made. We will attempt to consider the entire territory over which the discussion of mapping arises, but no doubt we will miss some of the critical distinctions and issues.

We have divided the paper into three main sections. The first section distinguishes the different kinds of entities that are related by analogy and similarity mappings, and some of their more salient properties. The second section discusses the different contexts or tasks that give rise to mappings. The third section catalogues the set of issues we have identified in the literature, and identifies some of the different solutions proposed or possible for each issue. In a concluding section we briefly discuss the implications of this framework for research.

2. WHAT IS MAPPED

The hypothesis we offer is that there are three fundamentally different kinds of entities that are mapped: systems, concepts, and properties and that all the other kinds of mappings discussed in the literature are variations on one of the three.

System Mapping. The mapping from the solar system to the atom that Gentner (Gentner, 1983) discusses is the classic example of a system mapping. In a system mapping it is critical to determine two types of mappings (Gentner, this volume):

1. Which components (i.e., concepts) in the source domain are mapped into which components in the target domain.
2. Which properties of each component (including relations between components) in the source domain are mapped into which properties in the target domain.

In the solar system/atom analogy, one first has to decide what components map (sun --> nucleus, planets --> electrons) and then what properties map (planets orbit the sun --> electrons orbit the nucleus).

Concept mapping. To answer the question (Collins, 1978) "Was Nixon a crook?" or to decide how likely Linda is to be a feminist bank teller (Tversky and Kahneman, 1980, Smith & Osherson, this volume) requires only a mapping across the properties of two concepts. There is no decomposition into components, as there is with a system mapping. So, in the case of Linda in Smith and Osherson's (this volume) account, you consider the properties of salary, education, and politics in the mapping process, comparing Linda and feminist bank tellers with respect to these properties.

Property mapping. The simplest kind of mapping specifies a particular property of two concepts for comparison, as when one judges whether an object 3 inches in diameter is more similar to a quarter or a pizza (Rips, this volume). (This example is actually a double mapping, discussed later under three-element comparisons, between a 3 inch object and a quarter, and between a 3 inch object and a pizza - system and concept mappings can also involve double mappings.) Property mappings differ from concept mappings in that the concepts are compared with respect to a particular property rather than with respect to many properties.

The critical distinction between these three kinds of mappings is that the system mappings involve component (or object) mappings as well as property mappings, that concept mappings involve multiple property mappings, and that property mappings involve individual properties of two concepts. The distinction between system and concept mappings is not entirely straightforward. For example, one elementary text we studied (Collins, Gentner, and Rubin, 1981) explained the composition of the earth by analogy to a peach. There is the crust which is analogous to the skin, the mantle analogous to the fruit, and the core analogous to the pit. This may appear to be a concept mapping, since it is a comparison of the properties of two concepts. But in fact it is a system mapping, since it requires first decomposing the earth and peach into their components (i.e., the three layers), and then comparing the properties of each pair of components (e.g., the skin and crust are both very thin), and their relations to each other. Thus the distinction between a system mapping and a concept mapping rests upon whether there is a two-stage process of first mapping an organized set of components and then the properties of each component (i.e., a system mapping) or a single-stage process of mapping properties (i.e., a concept mapping).

To give a second example of a system mapping that may be difficult to recognize, one might hypothesize (Collins and Michalski, 1987) that a bird's pitch depends on the length of the bird's neck, which is why ducks quack and geese honk, and more generally why small birds sing and big birds squawk (Malt and Smith, 1984). This hypothesis might be generated by analogy to the fact that human pitch (e.g., children vs. adult voices) depends on the length of the windpipe. To make the inference about birds by analogy to humans requires mapping windpipe length onto neck length, and human pitch onto bird pitch. Because the analogy involves both a mapping between their components (e.g. windpipes and neck) and a mapping of some of their components' properties (relative length), it is a system mapping. In this case the property mapped (*e.g.*, "pitch is inversely related to length") is a relational property in Gentner's (this volume) terms or a mutual dependency in Collins and Michalski's (Collins and Michalski, 1987) terms.

There are a number of other kinds of mappings discussed in the literature which we think are special cases of these three kinds of mappings. We will briefly describe each.

Procedure mapping. VanLehn and Brown (VanLehn and Brown, 1980) discuss mapping between the addition and subtraction procedures we learn in school and different addition and subtraction procedures with Dienes blocks (which are wooden blocks in three denominations: units are small squares, tens are ten unit blocks long, and hundreds are ten by ten unit blocks). Similarly, Anderson and Thompson (this volume) describes mapping between the procedure for factorial and that for summorial. Mappings of procedures are essentially system mappings, where the components of the procedure must first be mapped (e.g., unit blocks onto the numbers in the right hand column, etc.), and the manipulations on those components are subsequently mapped like properties.

Problem mapping. Ross (this volume), Holyoak & Thagard (this volume), and Carbonell (Carbonell, 1986), among others, discuss mapping between a problem you are trying to solve and an earlier problem you have solved. This kind of mapping is frequently used in science texts where students solve new problems by referring back to the sample problems worked in the text. Gick and Holyoak (Gick and Holyoak, 1980, 1983; Holyoak and Thagard, this volume) discuss the analogy between a fortress problem, where an army must split up in small units to capture a fortress, and Duncker's (Duncker, 1945) ray problem, where a ray source must be split in order to kill a tumor without destroying healthy tissued around it. Problem mappings require mapping of components first (e.g. ray --> army units, tumor --> fortress), and so they are system mappings.

Story Mappings. Gentner and Landers (Gentner and Landers, 1985) and Ross (this volume) have studied mappings between stories. These again are simply system mappings, where it is necessary first to map the characters or objects from one story to the other and then the relations or events between these entities.

There are undoubtedly other kinds of mappings that are made, but we think they will all be variations of the three kinds of mappings we have identified.

3. CONTEXTS IN WHICH MAPPINGS OCCUR

Various tasks or real world demands require different kinds of reasoning when relating entities. Our taxonomy of contexts in which mappings occur consists of two dimensions. *type of task* and *number of entities compared*. The overall structure of the taxonomy of contexts is shown in Table 1.

These two dimensions, type of task and number of elements, define a space of possible contexts in which mappings are made. There may be some cells empty in the space, but most combinations are possible.

3.1 TYPE OF TASK

Type of task breaks down into three basic categories: comparative judgements, mappings, and conceptual combinations. We will briefly describe six different kinds of comparative judgements, and then two kinds of mappings. Last, we will briefly discuss conceptual combination. The comparative-judgment types are derived primarily from the Rips (this volume) and Linda Smith (this volume) papers. This may not be a complete list of comparison judgements, but it covers the types discussed in this volume.

A. Comparative Judgements

1. **Similarity judgement**. Judging how similar two entities are is a common task in psychological experiments (Tversky, 1977; Rips, this volume, Smith & Osherson, this volume, Barsalou, this volume). Smith and Osherson (this volume) and Collins and Michalski (Collins and Michalski, 1987) argue that similarity judgments affects the certainty of many inferences people make. Similarity judgments obviously can apply to pairs of systems, concepts, or properties.

2. **Typicality judgments**. Typicality has been studied in psychology since Rosch (Rosch, 1975), and plays much the same kind of role in plausible reasoning as similarity (Collins and Michalski, 1987). Rips (this volume) has shown convincingly that typicality and similarity judgment are not always made in the same way, so they must be distinguished in any theory. Like similarity, typicality applies to pairs of systems, concepts, or properties.

3 **Categorization judgments**. Rips (this volume) discusses the similarity theory

Table 1

Contexts in which mappings occur

I. Type of Task

1. Comparative judgments

a. Similarity judgments
b. Typicality Judgements
c. Categorization judgments
d. Identity judgments
e. Overlap judgments
f. Difference judgments

2. Mappings

a. Property mappings
b. Component mappings

3. Conceptual combinations

II. Number of Entities Compared

1. Two-element mappings
2. Three-element mappings
3. Four-element mappings

of categorization, which he rejects. In any case, categorization requires a comparison between properties of two entities, the thing to be categorized and the category. Categorization only applies to systems and concepts, not to single properties of concepts, except when they are treated as concepts in their own right.

4. **Identity judgments**. Linda Smith (this volume) raises the issue of making identity judgments between entities – that is comparing whether all their properties are the same. Of course, no two entities are ever exactly the same (e.g. her examples of identical elements are not quite the same darkness or shape), so it is necessary to learn what degree of variability of a property can be called the same. Identity judgements therefore depend on context.

5. **Overlap judgments**. None of the papers in this volume mention overlap judgments (e.g. whether therapists are psychiatrists), but logically if one includes categorization and identity judgments, then overlap and difference judgments must also be included. Evaluating a "some" statement (e.g. "Some women are doctors") requires making an overlap judgment (Meyer, 1970).

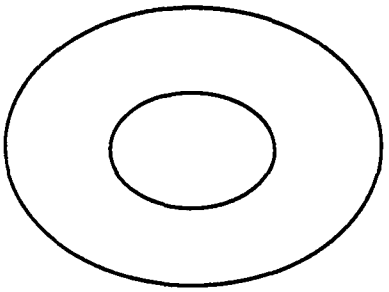
6. **Difference judgments**. The question of whether two entities are different (e.g. "Are whales fish?") also involves a comparison of properties. Like categorization, identity, and overlap judgments, difference judgments are contextually defined. For example, whales and fish are different, but both are animals and can be treated as the same in some contexts, such as grouping things as plants and animals.

The last four of these judgments: categorization, identity, overlap, and difference correspond to the four possible relations between two circles in Venn diagrams, as shown in Figure 1.

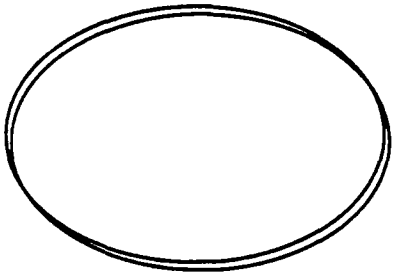
B. Mappings

The other type of task that is referred to frequently in the literature is one of mapping properties, components, or both from the source domain to the target domain.

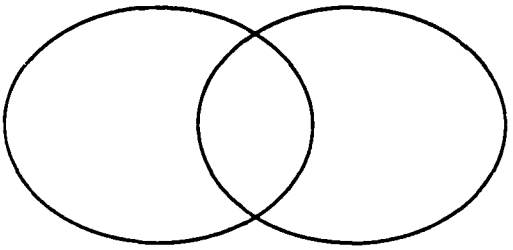
Property mapping. Most of the work on analogy (e.g. Anderson & Thompson, this volume. Gentner, this volume, Holyoak & Thagard, this volume) concerns itself with bringing properties (including relational properties) of objects in the source domain over into the target domain. A similarity or typicality judgment between the source and target is made before mapping a property over, and affects the certainty with which the property is believed to hold for the target domain. For example, before deciding that the pitch of birds depends on their neck length, based on an analogy to the human vocal tract, a person would compare humans and birds with respect to



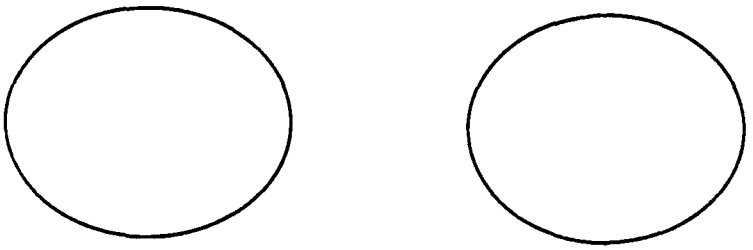
Categorization



Identity



Overlap



Difference

Figure 1 Four relations in comparing concepts.

their similarity, particularly on those properties related to sound production (in this case, properties of the relevant components, such as vocal cords and necks). A person's certainty about whether the property holds for birds depends on this similarity judgment.

Component mapping. Sometimes in the mapping of two systems, whole components are introduced by the mapping. In the earth/peach analogy, the text introduced two new components of the earth to students (the mantle and the core) in the course of explaining the analogy. This same thing can occur when people consider an analogy in their own mind (Collins and Gentner, 1980). For example, in relating the texture of foods to materials science, one might notice that chewiness corresponds roughly to elasticity, crispness to ductility, and then wonder what juiciness corresponds to. One possibility is liquid-filled porosity, a critical concept in geology.

C. Conceptual Combinations

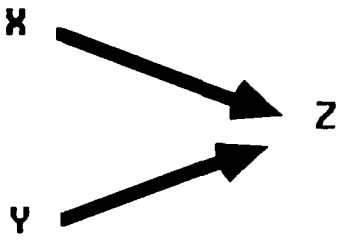
Smith and Osherson (this volume) raise the possibility that conceptual combination (feminist + bank teller --> feminist bank teller) is another task that a theory of mapping should address. We see conceptual combination, as they have modeled it, as primarily addressing the issue of how property mappings are combined when there is prior information about the properties involved in the target system. This becomes particularly important when learning or making predictions from multiple analogies, and in interpreting descriptive metaphors.

3.2 NUMBER OF ENTITIES COMPARED

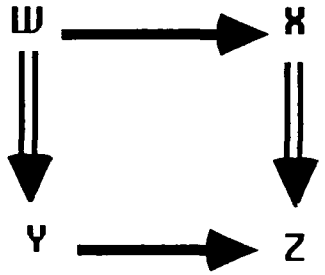
Number of entities compared is the other dimension we have identified with respect to the contexts in which analogies occur. This can range from two, as in the earth/peach mapping, to four as in analogies like wolf:dog::tiger:cat, and the geometric analogies considered by Evans (Evans, 1968). Slightly different constraints operate in two, three, and four-element mappings, shown in Figure 2.

Two-element comparisons. Many of the mappings discussed in the literature (e.g. Gentner's (this volume) water flow to heat flow mapping, and Holyoak & Thagard's (this volume) fortress problem to ray problem) are two-element mappings. In a two-element

Two-element comparison



Three-element comparison



Four-element comparison

Figure 2 The structure of different comparisons of concepts.

mapping there is a source from which properties or components are mapped onto a target. There are no other concepts, even implicitly, that are compared in a two-element comparison.

Three-element comparisons. A good example of a three-element mapping is the task used by Rips (this volume) where subjects had to decide whether a three inch object was more likely to be a pizza or a quarter. In many tasks that appear to be two-element mappings, there may be a comparison element implicit that subjects generate on their own in doing the mapping. For example, if you tell a child that a whale is a mammal, they may compare whether whales are more like mammals or fish, which is a three-element comparison. Three-element comparisons, therefore, compare properties of X to those of Y vs. Z.

Four-element comparisons. Standard analogy tests pose questions using the syntactic form W:X::Y:Z. We view such problems as falling into two categories, based on whether the analogy's interpretation depends on one or two comparisons. True four-element comparisons depend on both sets of comparisons, as in the analogy wolf:dog::tiger:cat. The within-group comparisons (e.g. wolf:dog and tiger:cat) determine the properties or dimensions along which the pairs differ (wildness or not), and the between group comparisons identify the dimensions along which the pairs are similar (feline or canine class membership). Evans (Evans, 1968) discusses the need for both kinds of comparisons (relating the components of both W and X and W and Y) in solving some geometric analogies.

Some analogies stated in the same syntactic fashion are more properly interpreted as analogies between two systems, where W and X are related in one system, while Y and Z are related in an analogous system. For example, Johnson-Laird (this volume) discusses, the analogy Water:Sluice::Gas:Jet. Here, there is a between-system mapping of water --> gas and sluice --> jet, but comparison of water and sluice, or gas and jet is not useful. Instead, there are relational systems relating each pair (e.g., a sluice directs water and a jet directs gas). True four-element mappings relate each concept in two different mappings, but in Johnson-Laird's example, a similarity judgment is required between the functional relations in the two systems.

4. ISSUES FOR A THEORY OF MAPPING

There are a number of issues running through the papers in this volume and the literature more generally. In part they reflect the set of subprocesses outlined by Gentner (this volume), but they have wider scope. Our attempt here is simply to delineate the set of issues as best we can, and to discuss possible resolutions to them. We start with the most microscopic issues and work up to the more macroscopic issues.

How are individual properties compared?

Potentially there are two kinds of properties that a theory must take into account: discrete properties (e.g. male or female) and continuous properties (e.g. size). Tversky and Gati (Tversky and Gati, 1982) have shown how it is possible to treat all continuous properties as if they were discrete. Another possibility is to treat all discrete properties as continuous (a person is on a continuum of male/female and most people fall near one or the other ends of the continuum).

Rips (this volume) addresses the question of how continuous properties are compared for different kinds of three-element comparisons: similarity, typicality, and categorization judgments, which he finds are judged differently. His results suggest that categorization judgments are based on the relative height of the distribution - e.g. a three inch object is more likely a pizza than a quarter, because the distribution of pizzas is higher at that point. His results for similarity judgments suggest both height of the distribution and distance from the mean (or mode) come into play. Typicality judgments appear to fall in between categorization and similarity, as if some subjects treat them like categorization judgments and others like similarity judgments (or perhaps they are combination judgments).

There are many possible functions for computing any of these judgments for example, similarity might be based on the relative distance between modes of the distribution compared, typicality judgments might be simply similarity judgments between a concept and its superconcept, as Smith & Osherson (this volume) assume. Rumelhart's (this volume) theory probably makes a prediction as to which of these functions will best fit the data, but he is not explicit on this point. Most of the other theories take no stand on this issue.

How are judgments from different properties combined?

Tversky (Tversky, 1977) proposes a combining function for similarity judgments, which Smith and Osherson (this volume) have adopted for their theory of decision making. The essence of the Tversky combination rule is that matching properties increase similarity and mismatching properties decrease similarity between concepts. Mismatching properties consist of two sets: property values of one concept that the other does not share, and property values of the other that the first does not share. Mismatching properties include properties where one concept has a known value and the other has no known value. Each of these three sets (one matching and two mismatching properties) is weighted appropriately depending on the direction of the judgment. Thus, people think North Korea is more like China than China is like North Korea, because there are many properties they know about China that do not apply to North Korea, but few properties they know about North Korea that do not apply to China (Tversky, 1977).

The Tversky rule is defined only over similarity judgments and discrete properties. If one adopts the view that all properties are continuous, then a modification of the Tversky rule is necessary. Whether it applies to other kinds of judgments (e.g. categorization judgments) is an open question. And, of course, there are an infinite number of other combination rules, some of which might still be viable given Tversky's (Tversky, 1977) data.

How do people access similar entities?

The question of access is fairly central to the papers of Ross, Gentner, Bransford et al., Barsalou, Brown and Kane, and Holyoak and Thagard (this volume). It is called "noticing" by Ross. All of these papers address the access issue for the case where the source must be found in memory. As Johnson-Laird (this volume) point out, the source is often given, as when a text explains that the earth is like a peach or the atom like a solar system. In Ross's paradigm, when one is working problems, a person may go back through a book to find a similar problem. This access may or may not be governed by the same properties as the access from memory.

Gentner (this volume) proposes that attributes (or superficial properties) govern access more than relational properties. This seems to accord fairly well with both her

data and those of Gick and Holyoak (Gick and Holyoak, 1983). Rumelhart (this volume) takes the position that access is governed by a match on all microfeatures, but in different contexts it may be attributes that match or higher-order relational features. These positions are compatible if one posits that superficial attributes are the most available, and therefore will usually dominate higher-order relations in most matches.

There is some evidence (e.g., (Chi, Feltovich and Glaser, 1981)) that part of becoming an expert is learning to pay attention to higher-order relations rather than superficial attributes. This also accords with Ross's (this volume) observation that superficial properties will mislead people if the principles underlying the problem (i.e. higher-order relations) are confusable. Brown and Kane (this volume) give evidence that functional fixedness and cognitive embeddedness of problem solving contexts are sources of diminished accessability to potential analogs in children, as well.

How is knowledge about the source reconstructed?

Ross (this volume) points out that people often have to reconstruct their knowledge about the source domain after they have accessed an analogy. This reconstruction process is guided by the knowledge being sought about the target domain. For example, if people are told that heat flow is like water flow (Gentner, this volume) since they do not have a particularly good understanding of water flow (Gentner and Gentner, 1983), they must in part figure out what they know about water flow: that it flows from one container to another as long as there is a difference in the height of the water in the two containers, that the surface area of the water in the container does not matter, that the flow rate is proportional to the diameter of the connection between the containers, etc. Which properties of the source domain people think of depends on what aspects of the target they are trying to understand, as Ross (this volume) has found in his studies.

What governs which properties are transferred?

This is the central argument animating most of the discussion in the analogy literature. We will briefly delineate the different positions.

Ortony (Ortony, 1979) advocates the position that *salience imbalance* governs transfer: that is, those properties are transferred that are important in the source domain but not important in the target domain. For example, Sam is a hippopotamus transfers fatness, since that is a typical property of hippos, but not of people.

Gentner (Gentner, 1983) proposes a syntactic theory that states that, in analogies, relational properties are transferred but attributes (i.e. non-relational properties) are left behind. Furthermore, according to her systematicity principle, relational properties that are a part of a system of relations (e.g. the large mass of the sun attracts the planets into orbiting around it) are more likely to be mapped across.

Holyoak and Thagard (this volume), Johnson-Laird (this volume), Carbonell (Carbonell, 1986), and Burstein (Burstein, 1986), while there are differences in their views, take a position on mapping that appears somewhat different from Gentner. Their position is that a system (or schema) of properties is mapped over, as Gentner proposes, but with two differences: (1) attributes will be mapped if they are part of the system, and (2) the major problem is to decide which system to map over. For example, if the analogy was made between the solar system and a person tanning themself under a sun lamp, the properties mapped would have to do with the heat being transmitted, the person rotating to cover all sides, the yellow color of the lamp, etc.

It turns out that the latter criticism may be handled by the structure mapping engine (Falkenhainer et al., 1986, Gentner, this volume) that was built recently to embody the Gentner theory. This system compares repesentations of two domains to decide which relations fit into a connected system that can be mapped into the target domain. Because it is effectively comparing all possible sets of relations between the objects considered, it is to some degree automatically chosing a "best system" to map. However, some pragmatic, contextual selection mechanisms will almost certainly be required as well. This is particularly true during learning, when people usually do not know enough about the target domain to pick out corresponding systems simply by matching (Burstein, 1986).

An important test of any of these computer models (Burstein, 1986, Carbonell, 1986, Gentner, this volume, Holyoak and Thagard, this volume) is whether they can select two different mappings from a source domain (e.g. the solar system) depending on what aspects of the source domain are relevant to the target domain (e.g. the atom vs. a person tanning). None of the models has, as yet addressed this central problem directly.

Whether goals and subgoals guide the selection of the system to be mapped often arises in the debate between these two positions. But that is probably because the latter researchers are all working with analogies in problem solving, whereas Gentner is dealing mainly with explanatory analogies. Certainly both sides would agree that goals are critical properties to map in problem-solving analogies and play the same central role that causal relations play in explanatory analogies.

Anderson and Thompson (this volume) rely on a set of three principles (i.e., "no function is content", "sufficiency of functional specification", and "maximal functional elaboration") to determine what is mapped. Although it is not clear to us exactly how these principles operate, they indicate the use of function as the main criteria for selecting what to map, and so would seem to fall into the latter camp.

In our view the positions of Gentner on the one hand and that of Holyoak and Thagard, Johnson-Laird, Carbonell, and Burstein on the other hand are not that far apart given the centrality of systems of properties or schemas that are mapped over. The Ortony theory is orthogonal to that issue, and could operate in conjunction with *some kind of system mapping*. Whether the Anderson and Thompson position is genuinely distinct, or reduces to the use of system properties as well, remains to be seen.

How are multiple mappings merged together?

This issue is raised by Burstein (Burstein, 1985, 1986, 1987), Spiro (this volume) and Collins and Gentner (Collins and Gentner, 1983). In Burstein's work, students were learning to program and were forced to combine the mappings of systems like puttings things in boxes and the the interpretation of arithmetic equalities in forming a mental model to understand computer statements like A=B+I. Collins and Gentner (Collins and Gentner, 1983, 1987) describe how subjects combined different analogies (e.g. billiard-ball analogy, a rocketship analogy, a crowded-room analogy) in understanding evaporation processes. It is clear that people frequently construct their understandings of systems by multiple mappings, and so theories will have to specify how conflicts are resolved about what properties to map from each analogy, and whether, in fact, some form of conceptual combination is required to merge related properties mapped from several different sources. In Burstein's model, conflicts between mappings are usually resolved by reasoning from specific examples in

the target domain that cause one or another analogical mapping to fail. However, the hypotheses that are eventually selected must still be integrated with what had been mapped previously or was otherwise known about the target domain (Burstein, 1987).

Burstein (Burstein, 1985) and Collins and Gentner (Collins and Gentner, 1983) also raise the issue of *vertical integration* of mental models. Analogies do not always map onto the same level of description of a target system. In such cases, one cannot directly merge analogs. Instead, the mapped structures must be maintained distinctly, and rules of correspondence formed between the different views or levels of abstraction described by the different analogical models.

How are mappings refined?

After a mapping is made, some properties carried over into the target domain will not apply. How are the correct properties identified and replaced? Both Burstein (Burstein, 1986) and Anderson and Thompson (this volume) address this question in the context of mapping computer program statements. In Burstein's model, analogically mapped predictions are compared to the actual results in target domain examples. If *the predictions are wrong, alternative structures are considered for mapping*, either from the same or a different source domain. Anderson and Thompson discuss several examples of failures due to overgeneralization from an analogy, and suggest that they may be handled by searching for contextual features that were *not* mapped, and adding them as preconditions.

Another kind of refinement occurs when successful analogies are extended to encompass new sets of corresponding systems or related causal principles. In addition to mapping new relational properties, this kind of analogical extension can lead to the introduction of new object or concept correspondences. For example, in the kinds of demonstration physics experiments that are often used to explain the diffraction and interference behavior of sound and light by using water wave tanks, a number of experimental objects are introduced to cause different wave behaviors. Each object that is introduced in these experiments must be related to an analogous object that causes a similar kind of interference with light or sound. In this sense, each new experiment described causes the refinement of the analogy between water waves and light or sound waves, because new objects and new causal implications are placed in parallel.

What is generalized from a mapping?

This is the question of how, when, and if generalizations are made based on a mapping between two domains. For example, one hypothesis might be that the corresponding components in the two systems are replaced by their common supersets, and the generalization is stored as a set of (possibly generalized) relations on these common supersets. Both Gentner (this volume), Anderson and Thompson (this volume) and Winston (Winston, 1982) have addressed this issue to some degree, although no specific claims have been made.

It is not at all clear that analogies always lead to new generalizations. Most analogies are only useful because they map one or two specific pieces of information from one domain to another. In such cases, the generation of a new general principle may not be warranted.

At the other extreme, attempting to generalize from an analogy that related radically different classes of objects by a new principle calls for a strong form of conceptual reclassification, as when sound and light are reclassified as waves. Very strong evidence of the analogy's pervasiveness may be needed for this kind of reclassification to occur. Alternatively, "bridging analogies" can be used to show why the analogy is justified. Clement (Clement, 1981, 1986) gives examples of series of bridging analogies designed to convince people of the generality of phyiscal laws. One set of these analogies shows how the behavior of a spring is related to the longitudinal and torsional flex of a wire, by considering intermediate cases where the wire is partially bent. Clement (Clement, 1986) also discusses Newton's analogy between the moon and an apple falling from a tree, with the a sequence of bridging analogs where a cannonball is fired at greater and greater speeds until it is in orbit around the earth.

How does the process of mapping develop?

This is the central issue raised by Linda Smith's paper (this volume). In it, she proposes that development proceeds from overall resemblance matches to identity matches and finally to dimensional matches. Her proposal perhaps is best summed up by saying that children learn to make finer discriminations in their comparison processing with age.

Her thesis raises the question of how children can make overall resemblance comparisons without being able to make individual property comparisons. This is not really a paradox from the vantage point of the kind of microfeature theory proposed by Rumelhart (this volume). Overall resemblance comparison in Rumelhart's theory can be carried out by comparing two concepts with respect to all their microfeatures. This requires no identification of microfeatures with particular properties (like color) of entities in the world. Based on the kind of perceptual learning described by Bransford and his colleagues (this volume), dimensions or subgroups of the microfeatures will emerge as contrastive sets of microfeatures that inhibit each other. Making an identity match would seem to require learning how much variability is possible on any dimension so that one can assess whether the difference between two entities falls below the normal range of variability on that dimension. In any case, the papers of Smith, Rumelhart, and Bransford et al. together promote a consistent picture of how similarity matching develops.

Are analogies helpful for learning?

This issue was raised by Halasz and Moran (Halasz and Moran, 1982). Their position is that if you give people explanatory analogies, such as the analogy that computer addresses are like boxes (Burstein, 1986) or that heat flow is like liquid flow (Gentner, this volume), you lead them to make more wrong mappings than helpful ones. So they argue that it is better to give people descriptions of the mechanisms involved, rather than analogies.

There are at least two arguments against the Halasz and Moran (Halasz and Moran, 1982) position. First, when people learn about novel systems, they are going to impute mechanisms to them. In order to understand any mechanistic description, they have to draw from their stock of basic mechanisms, such as Carbonell (this volume) or *Collins and Gentner* (Collins and Gentner, 1983) have described. So, whether you give students an analogy or not, they are going to make an analogy to some mechanism they already understand. The continuum from remembering, to reminding, to analogy that Rumelhart (this volume) describes is operating here. Subjects will pull in the mechanism they know about that matches most closely. By giving students an explicit analogy, you then accomplish two things. (a) you make sure they impute the best matching mechanism, and (b) you know what wrong inferences they are likely to draw, so that you can try to counter them as you explain the mechanism.

A second argument against the Halasz and Moran (Halasz and Moran, 1982) position is that the power of analogies for teaching derives from the fact that they provide a well-integrated structure that can be assimilated all at once. This structure may have acquired over a long period of time, as Vosniadou (this volume) shows for the solar system. So by telling someone the atom is like a solar system, the have a well-integrated structure acquired over many years that they can map as a whole in order to understand the atom. Thus they do not have to recapitulate the same long learning process for the atom. Analogies are particularly powerful where there is a competing structure already in place that the teacher is trying to dislodge.

The Halasz and Moran (Halasz and Moran, 1982) position, however, has to be correct if the analogy introduces too many wrong mappings. Therefore, we would argue that the issue is not whether analogies are helpful or harmful, but what determines when they are helpful vs. when they are harmful for learning.

5. CONCLUSION

Most researchers are working in a little corner of this framework, which is fine. One use of the framework is to help them see what the rest of the territory looks like in order to help them extend their theory to cover the whole territory. By trying to extend their theory in this way, it puts additional constraints on theory construction, which will help researchers refine their theories. Furthermore, as theories are extended to cover the whole domain, they will bump up against other theories in more ways which will lead to fruitful controversies and issues to be settled empirically. Psychology and artificial intelligence have a tendency to construct task-based theories and need to enforce on their theorists the desirability of constructing more global theories.

References

Burstein, Mark H. *Learning by Reasoning from Multiple Analogies.* Doctoral dissertation, Yale University, 1985.

Burstein, Mark H. Concept Formation by Incremental Analogical Reasoning and Debugging. In Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Ed.), *Machine Learning. Volume II.* Los Altos, CA: Morgan Kaufmann Publishers, Inc., 1986. Also appeared in the *Proceedings of the Second International Machine Learning Workshop*, Champaign-Urbana, IL., 1983.

Burstein, Mark H. Incremental Learning from Multiple Analogies. In *Proceedings of Analogica-85.* Boston, MA: Pitman, 1987. Forthcoming.

Carbonell, Jaime G. Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Ed.), *Machine Learning. Volume II.* Los Altos, CA: Morgan Kaufman Publishers, Inc., 1986.

Chi, M., Feltovich, P., and Glaser, R. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 1981, *5*(2), 121-152.

Clement, J. Analogy generation in scientific problem solving. In *Proceedings of the Third Annual Conference of the Cognitive Science Society.* Berkeley, CA: University of California, 1981.

Clement, J. Methods for evaluating the validity of hypothesized analogies. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Amherst, MA: University of Massachusetts, 1986.

Collins, Allan. Fragments of a Theory of Human Plausible Reasoning. In D. L. Waltz (Ed.), *Theoretical Issues in Natural Language Processing.* Urbana-Champaign, IL: University of Illinois, 1978.

Collins, Allan and Gentner, Dedre. A Framework for a Cognitive Theory of Writing. In L. W. Gregg and E. Steinberg (Eds.), *Cognitive processes in writing. An interdisciplinary approach.* Hillsdale, NJ: Erlbaum, 1980.

Collins, Allan and Gentner, Dedre. Multiple Models of Evaporation Processes. In *Proceedings of the Fifth Annual Conference of the Cognitive Science Society.* Rochester, NY: Cognitive Science Society, 1983.

Collins, Allan and Gentner, Dedre. How People Construct Mental Models. In N. Quinn and D. Holland (Eds.), *Cultural Models in Thought and Language.* Cambridge, UK: Cambridge University Press, 1987. In press.

Collins, A. and Michalski, R. The Logic of Plausible Reasoning: A Core Theory. Submitted to *Cognitive Science.*

Collins, A., Gentner, D. and Rubin, A. *Teaching Study Strategies* (Tech. Rep. Report No. 4794). Bolt Beranek and Newman Inc., 1981.

Duncker, K. On problem solving. *Phychological Monographs*, 1945, Vol. 58(270).

Evans, Thomas G. A Program for the Solution of Geometric Analogy Intelligence Test Questions. In Marvin L. Minsky (Ed.), *Semantic Information Processing*. Cambridge, Massachusetts: M.I.T. Press, 1968.

Falkenhainer, B., Forbus, K. and Gentner D. The Structure-Mapping Engine. In *Proceedings of AAAI-86*. Los Altos, CA. Morgan Kaufman, 1986.

Gentner, Dedre. Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, 1983, 7(2), 155-170.

Gentner, D. and Gentner, D. R. Flowing waters or teeming crowds: Mental models of electricity. In Gentner, D. and Stevens, A. L. (Eds.), *Mental Models*. Hillsdale, New Jersy. Lawrence Erlbaum Associates, 1983.

Gentner, D. and Landers, R. Analogical reminding: A good match is herd to find. In *Proceedings of the International Conference on Systems, Man and Cybernetics*. Tucson, AZ: University of Arizona, 1985.

Gick. Analogical problem solving. *Cognitive Psychology*, 1980(12), pp. 306-355.

Gick Schema induction and analogical transfer. *Cognitive Psychology*, 1983(15), pp 1-38.

Halasz, Frank and Moran, Thomas P. Analogy Considered Harmful. In *Proceedings of the Human Factors in Computer Systems Conference*. Gaithersburg, MD. , 1982.

Malt, Barbara C. and Smith, Edward E. Correlated properties in natural categories *Journal of Verbal Learning and Verbal Behavior*, 1984, *23*, 250-269.

Meyer, David E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, *1*, 242-300.

Ortony, Andrew. Beyond literal similarity. *Psychological Review*, 1979, *87*, 161-180.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology. General*, 1975, *104*, 192-233.

Tversky, A. Features of similarity *Psychological Review*, 1977, *84*, 327-352

Tversky, A. and Gati, I. Similarity, separability, and the triangle inequality. *Psychological Review*, 1982, *89*, 123-154.

Tversky, A. and Kahneman, D. Causal schemas in judgments under uncertainty. In M. Fishbein (Eds.), *Progress in social psychology*. Hillsdale, NJ. Erlbaum, 1980

VanLehn, K. and Brown, J. S. Planning Nets. A representation for formalizing analogies and semantic models of procedural skills. In Snow, R. E., Federico, P. and Montague, W. E. (Eds.), *Aptitude, Learning and Instruction. Volume 2*. Hillsdale, NJ. Erlbaum, 1980.

Winston, P. H Learning new principles from precedents and exercises. *Artificial Intelligence*, 1982, *19*, 321-350.

END

DATE

FILMED

8-88

DTIC